

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 17:34:32 ; Search time 3055.3 Seconds
(without alignments)
3801.338 Million cell updates/sec

Title: US-09-727-769a-5
555
Sequence: 1 ttggcgagtgtaattcctcgtg.....atgatccagctgtgattt 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Genembi:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_ov:*
- 7: gb_ov:*
- 8: gb_ov:*
- 9: gb_ov:*
- 10: gb_ov:*
- 11: gb_ov:*
- 12: gb_ov:*
- 13: gb_ov:*
- 14: gb_ov:*
- 15: gb_ov:*
- 16: gb_ov:*
- 17: gb_ov:*
- 18: gb_ov:*
- 19: gb_ov:*
- 20: gb_ov:*
- 21: gb_ov:*
- 22: gb_ov:*
- 23: gb_ov:*
- 24: gb_ov:*
- 25: gb_ov:*
- 26: gb_ov:*
- 27: gb_ov:*
- 28: gb_ov:*
- 29: gb_ov:*
- 30: gb_ov:*
- 31: gb_ov:*
- 32: gb_ov:*
- 33: gb_ov:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

1	555	100.0	555	6	AX113621	AX113621 Sequence
2	555	100.0	1080	6	AX113623	AX113623 Sequence
3	555	100.0	1380	6	AB046594	AB046594 Chryseobacterium
4	344.2	62.0	555	6	ARI59957	ARI59957 Sequence
5	344.2	62.0	555	6	ARI59957	ARI59957 Sequence
6	344.2	62.0	1080	6	ARI59962	ARI59962 Sequence
7	344.2	62.0	1080	6	ARI59962	ARI59962 Sequence
8	43	7.7	158254	2	AF429315	AF429315 Novel prote
9	41	7.4	125020	9	AF429315	AF429315 Novel prote
10	39.8	7.2	148000	2	AC034081	AC034081 Homo sapi
11	38.4	6.9	196932	2	AC100728	AC100728 Homo sapi
12	38	6.8	27068	3	CEC15A7	CEC15A7 Homo sapi
13	38	6.8	185947	3	AC092630	AC092630 Homo sapi
14	37.8	6.8	161729	5	AL590155	AL590155 Zebrafish
15	37.6	6.8	161221	5	AC022390	AC022390 Homo sapi
16	37.4	6.7	58282	2	AC106326	AC106326 Rattus no
17	37.2	6.7	50836	3	U67948	U67948 Caenorhabd
18	37.2	6.7	132749	2	AC006778	AC006778 Caenorhab
19	37.2	6.7	165373	2	AC093916	AC093916 Homo sapi
20	37	6.7	155344	2	AC026407	AC026407 Homo sapi
21	37	6.7	181946	2	AC097330	AC097330 Homo sapi
22	37	6.7	182974	2	AC105148	AC105148 Homo sapi
23	36.8	6.6	162371	9	AC016687	AC016687 Homo sapi
24	36.8	6.6	173201	9	AC008169	AC008169 Homo sapi
25	36.6	6.6	95566	9	AL161714	AL161714 Human DNA
26	36.6	6.6	154457	2	AC099239	AC099239 Rattus no
27	36.6	6.6	155842	2	AC025650	AC025650 Homo sapi
28	36.6	6.6	161479	8	AC090482	AC090482 Genomic s
29	36.6	6.6	177829	2	AC096872	AC096872 Rattus no
30	36.6	6.6	252360	2	AC091272	AC091272 Mus muscu
31	36.4	6.6	132977	2	RN510D20	RN510D20 Rattus no
32	36.4	6.6	204122	2	AL671867	AL671867 Mus muscu
33	36.2	6.5	1725	10	AB040609	AB040609 Rattus no
34	36.2	6.5	172792	9	AC062031	AC062031 Homo sapi
35	36.2	6.5	187445	30	AC024639	AC024639 Homo sapi
36	36	6.5	19576	6	AX251958	AX251958 Sequence
37	36	6.5	19576	6	AX349007	AX349007 Sequence
38	36	6.5	37170	9	AL356054	AL356054 Human DNA
39	36	6.5	68822	2	AC109141	AC109141 Mus muscu
40	36	6.5	68822	2	AC109141	AC109141 Mus muscu
41	36	6.5	179598	3	CNS01DRO	CNS01DRO Human chr
42	35.8	6.5	23342	9	CBR626J16	CBR626J16 Caenorhab
43	35.6	6.4	2489	1	AF112369	AF112369 Enrlichia
44	35.6	6.4	102126	2	AC098049	AC098049 Rattus no
45	35.6	6.4	170948	2	AC016931	AC016931 Homo sapi

ALIGNMENTS

RESULT 1
AX113621
LOCUS AX113621 555 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence AX113621
ACCESSION AX113621
VERSION AX113621.1 GI:13939799

SOURCE
ORGANISM Chryseobacterium sp. No. 9670.
Chryseobacterium sp. No. 9670.
Bacteria: CFB group; Flavobacteriia; Flavobacteriaceae;
Chryseobacterium.

REFERENCE
AUTHORS Yamaguchi, S.
TITLE Protein-deamidating enzyme, microorganism producing the same, gene encoding the same, production process therefor, and use thereof
JOURNAL Patent: EP 1106696-A 5 13-JUN-2001;
Amano Enzyme Inc. (JPN)

FEATURES
source location/Qualifiers
1..555 /organism="Chryseobacterium sp. No. 9670"
/db_xref="taxon:161487"

BASE COUNT 166 a 119 c 107 g 163 t
ORIGIN

Query Match 100.0%; Score 555; DB 6; Length 555;
Best Local Similarity 100.0%; Pred. No. 6,7e-138;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttggcgaagtgaatcttcgtatgtagcatcaatcttattcaatcaataaagaat 60
DB 1 TTGGGAGTGAATTCCTGATGTACATCAATTAATCTTTATTCATCAATAAAGAAAT 60
|||
QY 61 cagcttcggtacgtcgaagcggtccctacatcatcaatcaatcaatcaatcaat 120
DB 61 CAGCTTCGGTACCTCTACGGCGCTCCACCATCATCATTCAGATATCCTGTAGAC 120
|||
QY 121 ggaagttaagcaagccataaagaatgagacaactcaatgaacaagcgctatgct 180
DB 121 GGATGTTATGCAAGGCCCATATAGATGACAAATCTTATCAACACGGCTATACGT 180
|||
QY 181 gaaaaacaattgtatcacgaaacctaaaggcatcaacaggaacttgctgtgtg 240
DB 181 GAAAAACAATTTGTATACGAAACCTAAAGGCATCAACAGAACTTGTGTGCGCTGG 240
|||
QY 241 agctacccggttcgaatatgtgaagctataaataatgcttcgaggtgaagaaaga 300
DB 241 AGCTACCAAGTTGCAATTTGTTAGCTATATAAATGCTTCGGAGTAAAGGAAAAAGA 300
|||
QY 301 attatgaatccttcaactatctcaagcggtcctgtcaacagatagacagcaatg 360
DB 301 ATTATTTGATTCCTTCACTATTTTCAAGCGGCTCCTGTATACAGATACGATGGAAGCGT 360
|||
QY 361 tgcgttaacacctctgcgagatcgatccggttcccttattatgtaactgcaagaat 420
DB 361 TCGGTTAACACCTCTTGGGATCTGCAATCCGTTTCTTATGCTTAATCTGACAGAAAT 420
|||
QY 421 gttatcaagaagctcgaatcttcaactctgaatgaacaacatctataatcaaac 480
DB 421 GTTATTTACAGAGTCTCAGTAATTTCTTACCTGTATGACAAACATCTGATCAATCCAAC 480
|||
QY 481 tgtgactgaactaaattctcaactgcttccgagtgcttcccttcaactgacagatgta 540
DB 481 TGTGTACTGACTAAATTTTCACTGCTTCCGGATGTTCTCTTCAACGTGACCGGATGTA 540
|||
QY 541 tccagctgtgagatt 555
DB 541 TCCACCTGTGATTT 555
|||

RESULT 2
AX113623 1080 bp DNA linear PAT 30-MAY-2001
LOCUS AX113623
DEFINITION Sequence 7 from Patent EP1106696.
ACCESSION AX113623
VERSION AX113623.1 GI:13939800
KEYWORDS
SOURCE Chryseobacterium sp. No. 9670.
ORGANISM Chryseobacterium sp. No. 9670.
Bacteria; CFB group; Flavobacteriia; Flavobacteriaceae;
Chryseobacterium.
REFERENCE 1 (bases 1 to 1080)
Yamauchi, S.
TITLE Protein-deamidating enzyme, microorganism producing the same, gene
encoding the same, production process therefor, and use thereof
JOURNAL Patent: EP 1106696-A 7 13-JUN-2001;
Amano Enzyme Inc. (JP)
FEATURES
source location/Qualifiers
1..1080
/organism="Chryseobacterium sp. No. 9670"
/db_xref="taxon:161487"
61..1023
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC37857.1"

/db_xref="GI:13939801"
/translation="MKNLFLSMAPFVLPFENSCADNSNGOINKEKLSVNDKLD
GKTVPAGIDENEMIKVSEMLTPOFEYIKPTKEEYIGMLROAVKESVPHLEKP
NSNEIGVESASPEVDVRFKILRKVKGONKIASVDPVATINSI.FNOIKNSCGT
STASSPCTFRYPDPCCYARAHKROILMNGYCEKQFYGNLKAISGTCCVAMSH
VAILVSKNAGGVTEKRIIDPSLFSSGVDTARNNCVNTSGSASVSYANTAGNV
YRSPNSSYLDNMLINTNVLTKREFLSLSGCSPSPADVSSCGP"
466..1020
mat_peptide
/product="unnamed"

BASE COUNT 371 a 204 c 203 g 302 t
ORIGIN

Query Match 100.0%; Score 555; DB 6; Length 1080;
Best Local Similarity 100.0%; Pred. No. 6,6e-138;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttggcgaagtgaatcttcgtatgtagcatcaatcttattcaatcaataaagaat 60
DB 466 TTGGGAGTGAATTCCTGATGTACATCAATTAATCTTTATTCATCAATAAAGAAAT 525
|||
QY 61 cagcttcggtacgtcgaagcggtccctacatcatcaatcaatcaatcaatcaat 120
DB 61 CAGCTTCGGTACCTCTACGGCGCTCCACCATCATCATTCAGATATCCTGTAGAC 585
|||
QY 121 ggaagttaagcaagccataaagaatgagacaactcaatgaacaagcgctatgct 180
DB 121 GGATGTTATGCAAGGCCCATATAGATGACAAATCTTATCAACACGGCTATACGT 645
|||
QY 181 gaaaaacaattgtatcacgaaacctaaaggcatcaacaggaacttgctgtgtg 240
DB 181 GAAAAACAATTTGTATACGAAACCTAAAGGCATCAACAGAACTTGTGTGCGCTGG 705
|||
QY 241 agctacccggttcgaatatgtgaagctataaataatgcttcgaggtgaagaaaga 300
DB 241 AGCTACCAAGTTGCAATTTGTTAGCTATATAAATGCTTCGGAGTAAAGGAAAAAGA 765
|||
QY 301 attatgaatccttcaactatctcaagcggtcctgtcaacagatagacagcaatg 360
DB 301 ATTATTTGATTCCTTCACTATTTTCAAGCGGCTCCTGTATACAGATACGATGGAAGCGT 825
|||
QY 361 tgcgttaacacctctgcgagatcgatccggttcccttattatgtaactgcaagaat 420
DB 361 TCGGTTAACACCTCTTGGGATCTGCAATCCGTTTCTTATGCTTAATCTGACAGAAAT 885
|||
QY 421 gttatcaagaagctcgaatcttcaactctgaatgaacaacatctataatcaaac 480
DB 421 GTTATTTACAGAGTCTCAGTAATTTCTTACCTGTATGACAAACATCTGATCAATCCAAC 945
|||
QY 481 tgtgactgaactaaattctcaactgcttccgagtgcttcccttcaactgacagatgta 540
DB 481 TGTGTACTGACTAAATTTTCACTGCTTCCGGATGTTCTCTTCAACGTGACCGGATGTA 1005
|||
QY 541 tccagctgtgagatt 555
DB 1006 TCCACCTGTGATTT 1020
|||

RESULT 3
AB046594 1380 bp DNA linear BCT 27-APR-2001
LOCUS AB046594
DEFINITION Chryseobacterium proteolyticum prga gene for protein-glutaminase,
complete cds.
ACCESSION AB046594
VERSION AB046594.1 GI:12597204
KEYWORDS
SOURCE Chryseobacterium proteolyticum (strain:9670) DNA.
ORGANISM Chryseobacterium proteolyticum
Bacteria; CFB group; Flavobacteriia; Flavobacteriaceae;
Chryseobacterium.
REFERENCE 1 (sites)
Yamauchi, S., Jeenes, D.J. and Archer, D.B.
TITLE Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme

that deamidates glutamyl residues in proteins Purification,
characterization and gene cloning
Eur. J. Biochem. 268 (5), 1410-1421 (2001)
JOURNAL
MEDLINE
21153247
REFERENCE
2 (bases 1 to 1380)
AUTHORS
Yamaguchi, S.
Direct Submission
Submitted (26-JUL-2000) Shotaro Yamaguchi, Amano Pharmaceutical Co.
Ltd., Research and Development, Sui, Kagamigahara, Gifu 509-0108,
Japan (E-mail:LDV01447@nifty.ne.jp, Tel:81-583-79-1220,
Fax:81-583-79-1232)
JOURNAL

FEATURES

source

Location/Qualifiers
1..1380
/organism="Chryseobacterium proteolyticum"
/strain="9670"
/db_xref="taxon:118127"
181..243
/gene="prga"
181..1143
/gene="prga"
/codon_start=1
/transl_table=11
/product="protein-glutaminase"
/protein_id="BAB21508.1"
/db_xref="GI:12597205"
/translation="MKNLFLSMMAFVTVLTFPNSCADSNGNOEINGEKELSVNDSKLD
FKTVPVIGIDFENGMIKVSFMLETAQPELEIKPKENKOYICMLROAKNSPIHPIKRP
NSNEIKVESASPEVRYRTKILITKEVKQTNKLASVIFDVATLNLFOIKNOSGT
STASSPCITFRYPVDCVIAHMHMQILNNNSIDCEKQVYGLKASTGTCCVAAVSH
VALVSYKNASGVTETKRIIDPSLFFSSGPTDAMDACNVCNVCSSVSYANTAGNV
YRSPNSLYLYNNLINTNCLVLFKFSLSGSPSPADVSSCGF"
181..1143
/gene="prga"
586..1140
/gene="prga"
mat_peptide

BASE COUNT 472 a 247 c 247 g 414 t
ORIGIN

Query Match 100.0%; Score 555; DB 1; Length 1380;
Best Local Similarity 100.0%; Pred. No. 6.5e-138;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttggcgagtgtaattctctgtgtgagcttaacttaattcttattcaataaagaat 60
DB 586 TTGGCGAGTGAATTCCTGATGAGCTACATTAATTCTTATTCATCAATAAAGAAAT 645
QY 61 cagctctggagtaaccttaacgagctctcaatgcatgcatcagatatacctgtgagc 120
DB 646 CAGCTCTGGCGGTACCTTACGGCGCTCCTCACCATGCATTCACATTCACATTCCTGTAGAC 705
QY 121 ggaagtatgcaagagcccaataagatgagacaacttaatagaacaacgagctatgactgt 180
DB 706 GGATGTTATGCAAGAGCCCAATAGATGAGACAATCTTATGAACAACGGCTATGACTGT 765
QY 181 gaaaaacaattgtataaggaacctaagagcatcaacggaactgtgtgtgtgtgtgtg 240
DB 766 GAAAAACAATTTGTATACGGAACCTTAAGCATCAACAGAACTGCTGTGGCTGG 825
QY 241 agtaccaagctgcaaatgtgtaagatataaaaatgctccggagtaagaagaaaaaga 300
DB 826 AGCTACCACTTGCATATTGGTAGCTATTAATAATGCTTCCGAGTAGAGGAAAAAGAGA 885
QY 301 atattgactcctcaactatttcaacgagtcctgtlaacagatacagcatgagaagaacgct 360
DB 886 ATTATTGATCCTTACATATTTCGAAGGGGTCGTGTAACAGATACAGCATGAGAGAAAGCT 945
QY 361 tgggttaaacacctctggcgagatctgcatcgcttctctcttaatactgaagaat 420
DB 946 TGGTTAAACACCTTCTGGAGATCGCATCGTTTCCCTTATGCTATTAATGAGGAAAT 1005

QY 421 gttattacagaagctcagtaattcttacctgatatgacaacaactctgataataccac 480
DB 1006 GTTATTACAGAACTCCTAGTAATTCTTACTGTATGACAAACATTCGATCAATACCAAC 1065
QY 481 tgtgtactgactaaatttcaactgcttccggatgttctccttcaactgacagagatgta 540
DB 1066 TGTTACTGACTAAATTTTCACTCTTCCGAGTGTCTCTTCACTGACCGAGATGTA 1125
QY 541 tccagctgtgattt 555
DB 1126 TCCAGCTGTGGAATT 1140

RESULT 4
AR159957
LOCUS AR159957 555 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 5 from patent US 6251651.
ACCESSION AR159957
VERSION AR159957.1 GI:16222831
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 555)
AUTHORS Yamaguchi, S. and Matsunura, A.
TITLE Protein-deamidating enzyme, gene encoding the same, production
process thereof, and use thereof
Patent: US 6251651-A 5 26-JUN-2001;
JOURNAL Location/Qualifiers
FEATURES 1..555
source /organism="unknown"

BASE COUNT 159 a 122 c 112 g 162 t
ORIGIN

Query Match 62.0%; Score 344.2; DB 6; Length 555;
Best Local Similarity 76.7%; Pred. No. 1.7e-81;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 7 agtgtaatccctgtagtgaactaataattcttattcaataaagaatcagctc 66
DB 7 AGGTATTTCCTGATATCGGCAACCTGACAGATTATTATCCGATCAAAACACAGCT 66
QY 67 tggcgtaacctcagcgagctctcaatgcatcagatatacctgtgagagatgt 126
DB 67 TGGCGAATCTTACAGCATCTTCTCTGTATACCTTCAGATATCCGGTTGACGGATGT 126
QY 127 latgcaagagcccaataagatgagacaacttaatagaacaacgagctatgactgtgaaaaa 186
DB 127 TATGCAAGGGCTCACAATAATGAGACAATACTTATGAACGGCGCTATGACTGTGAAAAG 186
QY 187 caattgtatacgggaacctaagagcatcaacggaactgtgtgtgtgtgtgtgtgtgtgt 246
DB 187 CAGTTCGTATATGATATGATATGAGAGCTTCTTAACAGGAACATGCTGTATATGAGGTAT 246
QY 247 cagctgtcaatattgtgaactataaaaatgctccggagtaacggaagaaaaagaatatt 306
DB 247 CAGCTACCAATTTGGTAGCACTTCAAAAATGCTTCAAGAAATGTTGTAATAAAGATCATA 306
QY 307 gatccttcaactatttcaacgagtcctgtlaacagatacagcatgagaagaacgctgtgtc 366
DB 307 GATCCTTCATATTCTTCACAGCGGTCCGTGAACAGATTCGTGATGAGAGATGCGATGAC 366
QY 367 aacacctctggagatctgcatcggcttccctctatgataactgacgaagaatgtttat 426
DB 367 AACACAAGCTGCGAGATCTGCTGTATCTTCCATACCCCAATACAGCAGGAAAAATGTTTAC 426
QY 427 tacagaagctcagtaatttctcctgtatgagaacaacatactgataatactacactgtgta 486
DB 427 TACAGAAGCTCGTAGCTTATTAATGATGATACACATGATGATACCAATATGTGTA 486
QY 487 ctgactaaatttcaactgcttccggatgttctccttcaactgacacgagatgtatccagc 546
DB 487 CTGACTAAATTCTCACTGCTTCCGAGATCGCATCGTTTCCCTTATGCTATTAATGAGGAAAT 1005

Db 487 TTAACATATTCATCCCTTTCAGAGATGTCCTCTCCGCCAGACCAAGTTAGCAAGC 546
QY 547 tctgattc 555
|||||
Db 547 TGTGATT 555

RESULT 5
E59337 555 bp DNA linear PAT 07-FEB-2001
LOCUS Novel protein deamidation enzyme, gene encoding it, process for
DEFINITION producing the same, and utilization thereof.
ACCESSION E59337
VERSION E59337.1 GI:13023304
KEYWORDS JP 2000050887-A/3.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 555)
AUTHORS Shoharo,Y.A.M.M.
TITLE Novel protein deamidation enzyme, gene encoding it, process for
JOURNAL Patent: JP 2000050887-A 3 22-FEB-2000;
COMMENT AMANO PHARMACEUT CO LTD
OS Chryseobacterium gleum
PN JP 2000050887-A/3
PD 22-FEB-2000
PR 04-JUN-1999 JP 1999158703
PR SHOHARO YAMAGUCHI,AKIRA MATSUURA
PC C12N15/09,C12N9/80//A2D2/36,A23C9/154,A23C11/06,A23J3/00,PC
A23J3/10,
PC A23J3/16,A23J3/18,A23J3/34,A23L1/176,A23L1/23,A23L1/317,PC
(C12N9/80,C12R1:01),(C12N9/80,C12R1:20),C12N15/00 CC
FH Key Location/Qualifiers
FT source 1..555
location/Qualifiers
source 1..555
/organism='Chryseobacterium gleum'.
BASE COUNT 159 a 122 c 112 g 162 t
ORIGIN

Query Match 62.0%: Score 344.2; DB 6; Length 555;
Best Local Similarity 76.7%: Pred. No. 1.7e-81;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 7 agtgaattcctgtagctacataaattcttattcaataaataaagaatcagctc 66
|||||
Db 7 AGTGTATTCTCGATCTGCAACGCTGAACAGTTATTATCCACAGATCAAAAACAGGCT 66
|||||
QY 67 tgggtaccttaaggcgctcctcaacatgcatcaatcagatatccctgtagaaggatgt 126
|||||
Db 67 TGGGAACCTTCTACAGCATCTTCTCTGTATCACCTTCAATATCCGTTGACGAGT 126
|||||
QY 127 tatgaagaagccataaagatagacaataatcttaatgaaacaagcgctatgactgtaaaaa 186
|||||
Db 127 TATGCAAGGCTCACAAAATAGACAAATCTATTGAACGCCGCTTATGACTGTAAAAAG 186
|||||
QY 187 caattgtatacgyaaacctaaaggcatcaacaggaacttgctgtgcytggagctac 246
|||||
Db 187 CAGTTCGTAATATGTAATCTGAGACCTTCTACAGAACATGCTGTATCATGGGTATAT 246
|||||
QY 247 caagttgcaataatgtagtaagctataaataatgcttcggaagtaacggaanaagaattatc 306
|||||
Db 247 CACGTAGCAATTTTGGTAACTTCAAAAATGCTTCAGAGAAATGTTGAAAAAGAAATCATTA 306
|||||
QY 307 gatccttaacatttcaagcgctcgttaacagatacagcatgtagaagaacgcttgct 366
|||||
Db 307 GATCCTTATTAATTTCTCCAGCGGCTCTTAACAGATTCTGCATGAGAGCTGCATGTACC 366
|||||

QY 367 aacacctctgcgagatcgcatalccglttccctcttaatgttaactgacgaagaattttat 426
|||||
Db 367 AACACAAGCTGGATCTGCGTCTGTATCTTCATCGCAATATACAGAGAAATGTTTAC 426
|||||
QY 427 tacagaagctctagtaattcttacctgtatgacaacaatctgatatcaaccaacgtgtta 486
|||||
Db 427 TACAGAACTCCGTCAGGTTCATTAATGATATACAACTATGTAACCAATGTA 486
|||||
QY 487 ctgaataattcaactgcttcctgcgagatgcttcctcaacgtcagcggaattatcagc 546
|||||
Db 487 TTAACATATTCATCCCTTTCAGAGATGTCCTCTCCGCCAGACCAAGTTAGCAAGC 546
|||||

QY 547 tctgattc 555
|||||
Db 547 TGTGATT 555

RESULT 6
ARI59962 1080 bp DNA linear PAT 17-OCT-2001
LOCUS Sequence 11 from patent US 6251651.
DEFINITION ARI59962
ACCESSION ARI59962
VERSION ARI59962.1 GI:16222841
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Yamaguchi,S. and Matsuura,A.
TITLE Protein-deamidating enzyme, gene encoding the same, production
JOURNAL process thereof, and use thereof
PATENT: US 6251651-A 11 26-JUN-2001;
FEATURES location/Qualifiers
source 1..1080
/organism='unknown'
BASE COUNT 354 a 210 c 205 g 311 t
ORIGIN

Query Match 62.0%: Score 344.2; DB 6; Length 1080;
Best Local Similarity 76.7%: Pred. No. 1.6e-81;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 7 agtgaattcctgtagctacataaattcttattcaataaataaagaatcagctc 66
|||||
Db 469 AGTGTATTCTCGATCTGCAACGCTGAACAGTTATTATCCACAGATCAAAAACAGGCT 528
|||||
QY 67 tgggtaccttaaggcgctcctcaacatgcatcaatcagatatccctgtagaaggatgt 126
|||||
Db 529 TGGGAACCTTCTACAGCATCTTCTCTGTATCACCTTCAATATCCGTTGACGAGT 588
|||||
QY 127 tatgaagaagccataaagatagacaataatcttaatgaaacaagcgctatgactgtaaaaa 186
|||||
Db 127 TATGCAAGGCTCACAAAATAGACAAATCTATTGAACGCCGCTTATGACTGTAAAAAG 648
|||||
QY 187 caattgtatacgyaaacctaaaggcatcaacaggaacttgctgtgcytggagctac 246
|||||
Db 187 CAGTTCGTAATATGTAATCTGAGACCTTCTACAGAACATGCTGTATCATGGGTATAT 708
|||||
QY 247 caagttgcaataatgtagtaagctataaataatgcttcggaagtaacggaanaagaattatc 306
|||||
Db 247 CACGTAGCAATTTTGGTAACTTCAAAAATGCTTCAGAGAAATGTTGAAAAAGAAATCATTA 768
|||||
QY 307 gatccttaacatttcaagcgctcgttcaacagatacagcatgtagaagaacgcttgct 366
|||||
Db 769 GATCCTTATTAATTTCTCCAGCGGCTCTTAACAGATTCTGCATGAGAGCTGCATGTACC 828
|||||
QY 367 aacacctctgcgagatcgcatalccglttccctcttaatgttaactgacgaagaattttat 426
|||||
Db 829 AACACAAGCTGGATCTGCGTCTGTATCTTCATCGCAATATACAGAGAAATGTTTAC 888
|||||
QY 427 tacagaagctctagtaattcttacctgtatgacaacaatctgatatcaaccaacgtgtta 486
|||||


```

Db      889  TACAGAAGTCCTCAGGTTCTACTGTATGATATACAACTATGTGAAATTCATTTGTGA 948
QY      487  ctgactaaatttcacgtcttcgcgatgtcttcctcaccctcacgcagatgataccagc 546
Db      949  TTTAAACATATTTCTCATCCCTTTCAGGATGTTCTTCCACGACCACTGTAAGCAAGC 1008
QY      547  tctgagattt 555
Db      1009  TGTGATTT 1017

RESULT  7
LOCUS   E59342 1080 bp DNA linear PAT 07-FEB-2001
DEFINITION Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof.
ACCESSION E59342.1 GI:13023309
VERSION    E59342.1
KEYWORDS   JP 2000050887-A/8.
SOURCE     unidentified.
ORGANISM   unclassified.

REFERENCE 1 (bases 1 to 1080)
AUTHORS   Shoharo, Y.A.M.M.
TITLE      Novel protein deamidation enzyme, gene encoding it, process for
            producing the same, and utilization thereof
JOURNAL    Patent: JP 2000050887-A 8 22-FEB-2000;
            AMANO PHARMACEUT CO LTD
COMMENT    OS Chryseobacterium gleum
            PN JP 2000050887-A/8
            PD 22-FEB-2000
            PE 04-JUN-1999 JP 1999158703
            PR
            PI SHOTARO YAMAGUCHI, AKIRA MATSUURA
            PC C12N15/09,C12N9/80//A21D2/36,A23C9/154,A23C11/06,A23J3/00, PC
            A23J3/10,
            PC A23J3/16,A23J3/18,A23J3/34,A23J1/176,A23J1/23,A23J1/317, PC
            (C12N9/80,C12R1:01),(C12N9/80,C12R1:20),C12N15/00 CC
            FH Key Location/Qualifiers
            FT CDS (61)..(1020)
            FT mat_peptide (463)..(1017).
            Location/Qualifiers
FEATURES
source 1..1080
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT 354 a 210 c 205 g 311 t
ORIGIN

Query Match 62.0%; Score 344.2; DB 6; Length 1080;
Best Local Similarity 76.7%; Pred. No. 1,6e-81;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```

```

QY      307  gatccttcaactatttcaagcggctcctgtaacagatcacagcatgagaaacgcttcgctt 366
Db      769  GATCCTTCATATTTCTTCACAGGCTCCTGTACAGATTTCTGCAGAGAGCTGCATAC 828
QY      367  aacacctctgcggatctgcatccgttccctctatgtactaactcgaagaattgttat 426
Db      829  AACACAAGCTGCGGATCTCGCTGTATCTTCCACGCCAATACACAGGAAATGTTTAC 888
QY      427  tacagaagtcctgtaattcttaccgtatgatacaaaatctgataatccacttgta 486
Db      889  TACAGAAGTCCTCAGGTTCTACTGTATGATATACAACTATGTGAAATTCATTTGTGA 948
QY      487  ctgactaaatttcacgtcttcgcgatgtcttcctcaccctcacgcagatgataccagc 546
Db      949  TTTAAACATATTTCTCATCCCTTTCAGGATGTTCTTCCACGACCACTGTAAGCAAGC 1008
QY      547  tctgagattt 555
Db      1009  TGTGATTT 1017

RESULT  8
LOCUS   AC098659 158254 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-100L14, *** SEQUENCING IN PROGRESS
            *** 77 unordered pieces.
ACCESSION AC098659.3 GI:17973414
VERSION    AC098659
KEYWORDS   HTG: HTGS_PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 158254)
REFERENCE 1
AUTHORS   Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
            Alspbrooks, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barberia, J.,
            Benton, J., Binage, K., Blankenburg, K., Bonino, D., Bouck, J.,
            Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
            Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
            Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, R.,
            Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
            Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
            Davy-Carroll, L., Dederich, D.A., Delaney, R.R., Delgado, O.,
            Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Dreper, H.,
            Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
            Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
            Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
            Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,
            Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hayes, A.,
            Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
            Hollins, B., Homsl, F., Howard, S., Huber, J., Huily, S., Hume, J.,
            Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
            Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
            Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
            Louisged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
            Ma, J., Maheshwari, M., Kapua, P., Martin, R., Martindale, A.,
            Martinez, E., Massey, E., Mawlin, E., McLeod, M.P., Meador, M.,
            Mel, G., Metzger, M., Miner, G., Miner, T., Mitchell, T., Mohabbat, K.,
            Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
            Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenwo, S.,
            Oguh, M., Okunolu, G., Orzgunye, N., Oviado, R., Pace, A., Payton, B.,
            Peery, J., Peters, L., Peters, L., Pickens, R., Primus, E., Pu, J.L.,
            Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojucokan, I., Rolfe, H.,
            Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshari, N.,
            Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
            Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
            Tang, H., Tansey, J., Taylor, C., Taylor, T., Villalon, B., Thomas, N.,
            Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
            Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
            Watlington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S.,
            Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Weinstock, G. and Gibbs, R.
Direct Submission

2 (bases 1 to 158254)

Morley, K.C.

Direct Submission

Submitted (28-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064362.

COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GHVY

Center clone name: CH230-100L14

Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findPrimapList

Consensus quality: 121530 bases at least Q40

Consensus quality: 129199 bases at least Q30

Estimated insert size: 109535; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.5x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 77 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 5298: contig of 5298 bp in length

5299 5398: gap of unknown length

5399 9928: contig of 4530 bp in length

9929 10028: gap of unknown length

10029 14124: contig of 4096 bp in length

14125 14224: gap of unknown length

14225 18693: contig of 4469 bp in length

18694 18793: gap of unknown length

18794 22503: contig of 3710 bp in length

22504 22603: gap of unknown length

22604 24527: contig of 1924 bp in length

24528 24627: gap of unknown length

24628 28849: contig of 4222 bp in length

28850 28949: gap of unknown length

28950 31728: contig of 2779 bp in length

31729 31828: gap of unknown length

31829 35292: contig of 3464 bp in length

35293 35392: gap of unknown length

35393 38559: contig of 3167 bp in length

38560 38660: gap of unknown length

38661 40923: contig of 2264 bp in length

40924 41023: gap of unknown length

41024 43868: contig of 2845 bp in length

43869 43968: gap of unknown length

43969 46193: contig of 2225 bp in length

46194 46293: gap of unknown length

46294 48554: contig of 2261 bp in length

48555 48654: gap of unknown length

48655 51825: contig of 3171 bp in length

51826 51925: gap of unknown length

51926 54521: contig of 2596 bp in length

54522 54621: gap of unknown length

54622 56303: contig of 1682 bp in length

56304 56403: gap of unknown length

56404 58361: contig of 1958 bp in length

58362 58461: gap of unknown length

58462 61315: contig of 2854 bp in length

61316 61415: gap of unknown length

61416 63106: contig of 1891 bp in length

63107 63206: gap of unknown length

63207 65642: contig of 2436 bp in length

65643 65742: gap of unknown length

65743 67378: contig of 1636 bp in length

67379 69238: contig of 1760 bp in length

69239 69338: gap of unknown length

69340 71218: contig of 1880 bp in length

71219 71318: gap of unknown length

71320 73232: contig of 1914 bp in length

73233 75396: contig of 2064 bp in length

75397 75496: gap of unknown length

75497 77298: contig of 1802 bp in length

77299 77398: gap of unknown length

77399 80099: contig of 2601 bp in length

80100 81802: contig of 1703 bp in length

81803 81902: gap of unknown length

81903 83643: contig of 1741 bp in length

83644 83743: gap of unknown length

83744 85906: contig of 2163 bp in length

85907 87996: gap of unknown length

87997 88096: contig of 1990 bp in length

88097 89281: contig of 1185 bp in length

89282 89381: gap of unknown length

89382 91687: contig of 2306 bp in length

91688 93076: contig of 1289 bp in length

93077 93176: gap of unknown length

93177 95710: contig of 2534 bp in length

95711 95810: gap of unknown length

95811 98026: contig of 2216 bp in length

98027 98127: gap of unknown length

98128 99128: contig of 1002 bp in length

99129 101460: gap of unknown length

101461 101560: contig of 2232 bp in length

101561 103011: contig of 1451 bp in length

103012 103111: gap of unknown length

103112 105375: contig of 2264 bp in length

105376 105475: gap of unknown length

105476 107239: contig of 1764 bp in length

107240 107339: gap of unknown length

107340 108353: contig of 1016 bp in length

108354 108453: gap of unknown length

108454 110027: contig of 1572 bp in length

110028 110127: gap of unknown length

110128 111559: contig of 1432 bp in length

111560 111659: gap of unknown length

111660 113741: contig of 2082 bp in length

113742 113841: gap of unknown length

113842 115728: contig of 1887 bp in length

115729 115828: gap of unknown length

115829 116829: contig of 1001 bp in length

116830 116929: gap of unknown length

116930 11826: contig of 1897 bp in length

11827 118926: gap of unknown length

118927 120611: contig of 1685 bp in length

120612 120711: gap of unknown length

120712 121981: contig of 1270 bp in length

121982 122081: gap of unknown length

122082 123463: contig of 1382 bp in length

123464 123563: gap of unknown length

123564 124684: contig of 1121 bp in length

124685 124784: gap of unknown length

124785 125849: contig of 1065 bp in length

125850 125949: gap of unknown length

Query Match	Best Local Similarity	Matches	Score	DB	Length
125950	126974	contig of 1025 bp in length	7.7%	DB 2	Length 158254
126975	127074	gap of unknown length	Pred. No. 0.54		
127075	129000	contig of 1926 bp in length			
129001	129100	gap of unknown length			
316	ctatttcgaagggctcctcgtlaacagataaagatbgaagaagcttcgtcgtlaacacctct	375			
4289	ctcgcgtctcactctgtgcattttacaaagaaatttcgcatctacatctggttgaaacctcatt	4348			
376	tgcgagatctgcacgcggttccttccttaagctaaactatcgtcgaagaagatgttatcata	430			
4349	AAATCTCTGTAAACCGTTTCTGTTATGGAAGAAACTAGGAATGTCATATTACA	4403			
RESULT 9	AF429315/c	125020 bp	DNA	linear	PRI 18-JAN-2002
LOCUS	AF429315				
DEFINITION	Homo sapiens junctophilin 3 (JPH3) gene, partial cds.				
ACCESSION	AF429315				
VERSION	AF429315.1	GI:17646244			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 125020) Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S., Ingersoll,Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A., Potter,N.T., Ross,C.A. and Margolis,R.L.				
TITLE	A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2				
JOURNAL	Nat. Genet. 29 (4), 377-378 (2001)				
MEDLINE	21583737				
PUBMED	11694876				
REFERENCE	2 (bases 1 to 125020) Holmes,S.E., Ingersoll,Ashworth,R.G., Ross,C.A. and Margolis,R.L.				
AUTHORS	Direct Submission				
JOURNAL	Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA				
FEATURES	Location/Qualifiers				
source	1..125020 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="16q24.3: between D16S120 and WI-12410" /note="Isolated from a patient with Huntington's disease-like 2 (HDL2)" Disease-Like 2 (HDL2)" complement(35581..35746) /rpt_type=" tandem /rpt_unit="ctg complement(<35507..>36887) /gene="JPH3" /product="junctophilin 3" complement(<35507..>36887) /gene="JPH3" /note="JPH3" complement(<36507..36887) /gene="JPH3" /note="component of the junctional complex between plasma membrane and endoplasmic reticulum" /codon_start=1 /product="junctophilin 3" /protein_id="AAL04941.1" /db_xref="GI:17646245" /translation="MSSGGRFPDQSGSYCGGMDGRAHGHGVTGPGGSEYTGSMW HGFEVLGYTTWPSNLTGYGTAAQGRMRHIGLESQKVVYKKEWTHGFGRIYVRCACG NGAIEGWSMGLDDGTGTETYSOG"				
CDS					
BASE COUNT	29056 a 32731 c 30696 g 28283 t 4254 others				
ORIGIN					

```

Query Match Similarity 7.4%; Score 41; DB 9; Length 125020;
Best Local Similarity 12.8%; Pred. No. 1.9; Mismatches 165; Indels 0; Gaps 0;
Matches 47; Conservative 154;

Oy 130 gcaagagccataagatgagacaatcttaatgacaacagctatgactgtgaaacaa 189
Db 66546 GSSMKAMCICAMMMWMMWRMAAEMAAGAAGCGMGRMMRPYMYRRPMKGRSSRYWA 66487
Oy 150 ttgtatcaggaacacctaaaggcatcaacaggaactctgtgtgctgtgagcaccac 249
Db 66486 MYAGWGRRRGRCYRAYMYWMSRRATSKAMGAWGMYYTRPMKKYTKTKTKTGTGYCWC 66427
Oy 250 gtgcgaatcttgtaagcagataaaatgcttcgcgagtaacgaaaaaagattatgat 309
Db 66426 GYMMMKRMAAATYMTTYYTYAMRRKKYTKKAYKARWCMKCKMMWMMKKAWSMMWAMC 66367
Oy 310 ccttaactatttcaagcggctcgtgtaacagatatacagatgagagaacgcttgctaac 369
Db 66366 YGGYWAASMTNCASKASAYYRKGYSKRYMMKSYRSOSWSTYYTARARMMARMRMRMRGR 66307
Oy 370 acctcttcgcgagtcgcagtcgccttccttccttcgtctaatacgcagagaatttatatc 429
Db 66306 ATSRKKCKMKMKMAGCKMTYMAARMCMCGSSRRRSAAASRMAGSKRRMSSTYYCYCAC 66247
Oy 430 agaagctcctaagattcttaccctgtagtatacaacaactgatacatatacaactgttact 489
Db 66246 AMAAAMAMWMAAAATYRGYCKMSRMGMAMRMAMRMAMMMWRMRAKCMHGRYKMCYRGCS 66187
Oy 490 actaaa 495
Db 66186 ASRMWR 66181

RESULT 10
AC094081/c 148000 bp DNA linear HTG 14-SEP-2001
LOCUS Homo sapiens chromosome 5 clone CTC-205D2, LOW-PASS SEQUENCE
DEFINITION
AC094081
AC094081
AC094081.1 GI:15617642
VERSION HTG: HTGS_PHASE0.
KEYWORDS
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 148000)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 148000)
DOE Joint Genome Institute.
Direct Submission
Submitted (14-SEP-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 263973
Center clone name: CIT-HSPC_205D2
-----
Summary Statistics
Consensus quality: 109929 bases at least Q40
Consensus quality: 130319 bases at least Q30
Consensus quality: 134832 bases at least Q20
Estimated insert size: 160000; pulse field gel estimation
Estimated insert size: 143600; sum-of-contigs estimation
Quality coverage: 2.79 in Q20 bases; pulse field gel estimation
Quality coverage: 3.11 in Q20 bases; sum-of-contigs estimation.

```

* NOTE: This record contains 45 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 1345: contig of 1345 bp in length
 * 1346 1445: gap of unknown length
 * 1446 2511: contig of 1066 bp in length
 * 2512 2611: gap of unknown length
 * 2612 3875: contig of 1264 bp in length
 * 3876 3975: gap of unknown length
 * 3976 5634: contig of 1659 bp in length
 * 5635 5735: gap of unknown length
 * 5735 6917: gap of unknown length
 * 6917 8138: contig of 1122 bp in length
 * 8139 8238: gap of unknown length
 * 8239 9444: contig of 1206 bp in length
 * 9445 9544: gap of unknown length
 * 9545 10707: contig of 1163 bp in length
 * 10708 10807: gap of unknown length
 * 10808 11935: contig of 1127 bp in length
 * 11935 12034: gap of unknown length
 * 12035 13169: contig of 1135 bp in length
 * 13170 13270: gap of unknown length
 * 13270 14619: contig of 1350 bp in length
 * 14620 14719: gap of unknown length
 * 14720 16497: contig of 1778 bp in length
 * 16498 16598: gap of unknown length
 * 16598 18234: contig of 1637 bp in length
 * 18235 18335: gap of unknown length
 * 18335 20429: contig of 2095 bp in length
 * 20430 20529: gap of unknown length
 * 20530 22800: contig of 2271 bp in length
 * 22801 22900: gap of unknown length
 * 22901 25553: contig of 2653 bp in length
 * 25554 27672: gap of unknown length
 * 27673 27772: gap of unknown length
 * 27773 29756: contig of 1984 bp in length
 * 29757 29856: gap of unknown length
 * 29857 31387: contig of 1530 bp in length
 * 31387 31487: gap of unknown length
 * 31487 33887: contig of 2400 bp in length
 * 33887 33986: gap of unknown length
 * 33987 35894: contig of 1908 bp in length
 * 35895 35994: gap of unknown length
 * 35995 38807: contig of 2813 bp in length
 * 38808 38908: gap of unknown length
 * 38908 41243: contig of 2336 bp in length
 * 41244 41343: gap of unknown length
 * 41344 43289: contig of 1946 bp in length
 * 43290 43390: gap of unknown length
 * 43390 45626: contig of 2237 bp in length
 * 45627 47930: gap of unknown length
 * 47931 48030: gap of unknown length
 * 48031 51419: contig of 3388 bp in length
 * 51419 55433: gap of unknown length
 * 55434 55533: contig of 3915 bp in length
 * 55534 58444: gap of unknown length
 * 58445 58545: gap of unknown length
 * 58545 62187: contig of 3643 bp in length
 * 62188 65288: gap of unknown length
 * 65288 65729: contig of 3441 bp in length
 * 65729 65828: gap of unknown length

* 65829 68678: contig of 2850 bp in length
 * 68679 68778: gap of unknown length
 * 68779 71850: contig of 3072 bp in length
 * 71851 71950: gap of unknown length
 * 71951 75194: contig of 3244 bp in length
 * 75195 75294: gap of unknown length
 * 75295 78698: contig of 3404 bp in length
 * 78699 82388: contig of 3590 bp in length
 * 82389 82488: gap of unknown length
 * 82489 86689: contig of 4201 bp in length
 * 86690 86789: gap of unknown length
 * 86790 92232: contig of 3443 bp in length
 * 92233 92332: gap of unknown length
 * 92333 97989: contig of 5657 bp in length
 * 97990 98089: gap of unknown length
 * 98090 103185: contig of 5095 bp in length
 * 103185 103285: gap of unknown length
 * 103285 109565: contig of 6281 bp in length
 * 109566 109665: gap of unknown length
 * 109666 119541: contig of 9876 bp in length
 * 119542 119641: gap of unknown length
 * 119642 126139: contig of 6498 bp in length
 * 126140 126239: gap of unknown length
 * 126240 132394: contig of 6155 bp in length
 * 132395 132494: gap of unknown length
 * 132495 148000: contig of 15506 bp in length.
 * 148000
 * location/Qualifiers
 * 1. 148000
 * /organism="Homo sapiens"
 * /db_xref="taxon:9606"
 * /chromosome="5"
 * /clone="CTC-205D2"
 * /clone_lib="CalTech human BAC library C"
 * 39885 a 32077 c 31075 g 40373 t 4590 others

FEATURES

source

BASE COUNT 39885 a 32077 c 31075 g 40373 t 4590 others
 -ORIGIN

Query Match 7.2%; Score 39.8; DB 2; Length 148000;
 Best Local Similarity 49.3%; Pred. No. 3.9;
 Matches 104; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 18 tgatgactcatatatttcaatcaataagaatcgtctgcggtaccc 77
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 50593 TGGTTTGTAAACATACATCTATACAGATTGACATTTACTTGTAGCGCTC 50534
 OY 78 tacgcgcctccatcatcatcatcagatcctatgaagcggatgtatgcaagac 137
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 50533 TTTTGTCCAAACGATGCGGTTCCATACAGACGCTCTGGACACATGTAGTGATGAC 50474
 OY 138 ccataagatgagacaatcttaatagaacaacgctatgactgtgaaacaaattgtata 197
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 50473 AGATAAATATGCGTAGTCGCAAAATTAAGGTATGCTAAGTAAATGCAAAATGTGATT 50414
 OY 198 cggaaacctaaagcattcaacacaggaactgtc 228
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 50413 TCCAAACAGTATTCAAAAAAGAAATGC 50383

RESULT 11

AC100728 AC100728 196932 bp DNA linear HTG 22-NOV-2001
 LOCUS Mus musculus clone RP24-279H4, WORKING DRAFT SEQUENCE, 29 unordered
 DEFINITION pieces.
 AC100728
 AC100728.1 GI:17048095
 ACCESSION
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 196932)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE
JOURNAL
REFERENCE
AUTHORSMus musculus, clone RP24-279H4
Unpublished
2 (bases 1 to 196932)TITLE
JOURNAL
COMMENTSubmitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996.1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L17435

Center clone name: 279.H.4

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 185765 bases at least Q40

Consensus quality: 191126 bases at least Q30

Consensus quality: 193031 bases at least Q20

Insert size: 188000; agarose-IP

Insert size: 194132; sum-of-contigs

Quality coverage: 6.9 in Q20 bases; agarose-IP

Quality coverage: 6.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 29 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 7249: contig of 7249 bp in length
 * 7250 7349: gap of 100 bp
 * 7350 7942: contig of 593 bp in length
 * 7943 8042: gap of 100 bp
 * 8043 9463: contig of 1421 bp in length
 * 9464 9563: gap of 100 bp
 * 9564 10770: contig of 1207 bp in length
 * 10771 10870: gap of 100 bp
 * 10871 11894: contig of 1024 bp in length
 * 11895 11994: gap of 100 bp
 * 11995 13096: contig of 1102 bp in length
 * 13097 13196: gap of 100 bp
 * 13197 14617: contig of 1421 bp in length
 * 14618 14717: gap of 100 bp

14718 14939: contig of 222 bp in length
 * 14940 15039: gap of 100 bp
 * 15040 17365: contig of 2326 bp in length
 * 17366 17465: gap of 100 bp
 * 17466 20632: contig of 3167 bp in length
 * 20633 20732: gap of 100 bp
 * 20733 70675: contig of 49943 bp in length
 * 70676 70775: gap of 100 bp
 * 70776 71963: contig of 1188 bp in length
 * 71964 72063: gap of 100 bp
 * 72064 74943: contig of 2880 bp in length
 * 74944 75043: gap of 100 bp
 * 75044 78141: contig of 3098 bp in length
 * 78142 78241: gap of 100 bp
 * 78242 81595: contig of 3354 bp in length
 * 81596 81695: gap of 100 bp
 * 81696 84991: contig of 3296 bp in length
 * 84992 85091: gap of 100 bp
 * 85092 88215: contig of 3124 bp in length
 * 88216 88315: gap of 100 bp
 * 88316 91555: contig of 3240 bp in length
 * 91556 91655: gap of 100 bp
 * 91656 96337: contig of 4682 bp in length
 * 96338 96437: gap of 100 bp
 * 96438 100411: contig of 3974 bp in length
 * 100412 100511: gap of 100 bp
 * 100512 106473: contig of 5962 bp in length
 * 106474 106573: gap of 100 bp
 * 106574 113018: contig of 6445 bp in length
 * 113019 113118: gap of 100 bp
 * 113119 120051: contig of 6933 bp in length
 * 120052 120151: gap of 100 bp
 * 120152 128451: contig of 8300 bp in length
 * 128452 128551: gap of 100 bp
 * 128552 139502: contig of 10951 bp in length
 * 139503 139602: gap of 100 bp
 * 139603 150289: contig of 10687 bp in length
 * 150290 150389: gap of 100 bp
 * 150390 172565: contig of 22176 bp in length
 * 172566 172665: gap of 100 bp
 * 172666 195587: contig of 22922 bp in length
 * 195588 195687: gap of 100 bp
 * 195688 196932: contig of 1245 bp in length.

FEATURES

source

misc_feature
 1. 196932
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-279H4"
 /clone_lib="RPCI-24 Male Mouse BAC"
 1. 7249
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"
 7350..7942
 /note="assembly-fragment"
 8043..9463
 /note="assembly-fragment"
 9564..10770
 /note="assembly-fragment"
 10871..11894
 /note="assembly-fragment"
 11995..13096
 /note="assembly-fragment"
 13197..14617
 /note="assembly-fragment"
 14718..14939
 /note="assembly-fragment"
 15040..17365
 /note="assembly-fragment"
 17466..20632
 /note="assembly-fragment"
 20733..70675
 /note="assembly-fragment"

```

BASE COUNT      52753 a 45059 c 43133 g 53177 t      2810 others
ORIGIN
misc_feature    70776..71963
                /note="assembly-fragment"
misc_feature    72064..74943
                /note="assembly-fragment"
misc_feature    75044..78141
                /note="assembly-fragment"
misc_feature    78242..81595
                /note="assembly-fragment"
misc_feature    81696..84991
                /note="assembly-fragment"
misc_feature    85092..88215
                /note="assembly-fragment"
misc_feature    88316..91555
                /note="assembly-fragment"
misc_feature    91656..96337
                /note="assembly-fragment"
misc_feature    96438..100411
                /note="assembly-fragment"
misc_feature    100512..106473
                /note="assembly-fragment"
misc_feature    106574..113018
                /note="assembly-fragment"
misc_feature    113119..120051
                /note="assembly-fragment"
misc_feature    120152..128451
                /note="assembly-fragment"
misc_feature    128552..139502
                /note="assembly-fragment"
misc_feature    139603..150289
                /note="assembly-fragment"
misc_feature    150390..172565
                /note="assembly-fragment"
misc_feature    172666..195587
                /note="assembly-fragment"
misc_feature    195688..196932
                /note="assembly-fragment"
                clone_end:177
                vector_slide:right"

```

RESULT	12
CEC15A7/c	
LOCUS	CEC15A7
DEFINITION	CEC15A7
ACCESSION	27068 bp
VERSION	DNA
KEYWORDS	linear
SOURCE	INV 24-JAN-2002
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	

JOURNAL	MEDLINE	REMARK	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Investigating biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998)	99069613	The C.elegans Sequencing Consortium.	2 (bases 1 to 27068)	McMurray A.A.	Direct Submission	Submitted (07-NOV-1995) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: joesanger.ac.uk or rwematacoe.wustl.edu	Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information. Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is not the entire insert of clone C15A7. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone C15A7 is at 1 in this sequence. The true right end of clone C15A7 is at 16201 in sequence Z67738. The true left end of clone W03G11 is at 26969 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence Z67734. The end of this sequence (26969..27068) overlaps with the start of sequence Z67738. For a graphical representation of this sequence and its analysis see: http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C15A7 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small] overlap between neighbouring submissions.
FEATURES	SOURCE	1..27068	Location/Qualifiers				

```

/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="C15A7"
join(14093..14194,14238..14387,14874..15193,15238..15340,15524..15631,15753..15869,15935..16147,17039..17120,17265..17422,17639..17855,17902..18090,18369..18524,18569..18673,18943..19079,19126..19340)
/gene="C15A7.1"
join(14093..14194,14238..14387,14874..15193,15238..15340,15524..15631,15753..15869,15935..16147,17039..17120,17265..17422,17639..17855,17902..18090,18369..18524,18569..18673,18943..19079,19126..19340)
/gene="C15A7.1"
(note="contains similarity to Pfam domain: PF00065 (Neurotransmitter-gated ion-channel), Score=15.1, E-value=0.00034, N=2"
/codon_start=1
/protein_id="CAAG1530.1"
/db_xref="GI:3874299"
/translation="MIPMLMILVNTPTIRVSEADESESRQAVYDLHRIKSDVGP
LDGSMPLSPVIGKQALDHSRDEKPTLPIFVPMSEDLDDHRSKDFV
RTLEFVDFMLQIKTRPKRIYNTYEROSKTIIDLSTLVENDHSKGFVTTIKL
KATFEDVEKGYPHDFNCSSEFIPNNMADLFSVSMILKATKATKDFDTPVAVRD
FOINDFEDKYLPHNFRVLDLDESPRNLSPVSVFVTLQKRVHQLQAVP
ICTGILSGFSEFSEALVAFNLISEYLFSPNIDVLTDEDDGPTILAFITLAE
TMSITGMEFTITKOKAFVNLSEKPSDFGNIRRLRLIDKFLAMLYA
SEFGKPTFELDFPQRLAEMLTKLVKNQNNPVLAOSYKINSFSLSESYWS
VFCARVYSSCGVMGNGKILHGDDDHGMSVINAELVYQVPESTFQWV
DERLKMNPAVNSGVEYIEKTEFEKRDNNCAPIYKRSYDRRLSELDLSESDARTL
ISYELIKTALQITVNTTKQSEFGEFDPDYQCSMTLPNNQADRFVSPNCNPK
ENVEHRAVRVADHLDMCVESNIFPTAFVYTAELVYTGEMAKTEFPRMLNFRVN

```

[illegible]

Query Match 6.8%; Score 36; DB 3; Length 27068;
Best Local Similarity 48.2%; Pred. No. 12;
Matches 107; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Db	15755	CATCTGTAATGAATGCTTTAATAACCAACGTATCTCTTGTTTAAAAAAAATAAGGGA	15696
QY	199	ggaacctaagaagcatcaacacagaactctgctgtgctgctgagctacacgcttgcata	258
Db	15695	AATTACTTAAAAACCTATAAATAAAGTTACAGCCCAAGCTCTAGACCTCAAAATATTAAAC	15636
QY	259	ttggttaagcttaaaaatgctctccgagtaacaggaaaaaagaatatatgtacccctta	318
Db	15635	TTACGTCGCAATGAAACATCATTTGGAGCGTAAAGCTGAAGATGAACATTAACGTTTTCCT	15576
QY	319	tttcaagcggtccctgtacagatcacacatcgaggaacgct	360
Db	15575	TTGAAGAACTATTGTGAACCTGACAAAGATCGCAACAGCT	15534
RESULT	13		
LOCUS	AC092630	185947 bp	DNA linear PRI 08-FEB-2002
DEFINITION	Homo sapiens BAC clone RP11-321C18 from 2, complete sequence.		
ACCESSION	AC092630	AC032022	
VERSION	AC092630.3	GI:18497222	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 185947)		
JOURNAL	Sulston, J.E. and Waterston, R.		
MEDLINE	Toward a complete human genome sequence		
REFERENCE	Genome Res. 8 (11), 1097-1108 (1998)		
AUTHORS	2 (bases 1 to 185947)		
TITLE	Pearman, C., Abbott, A., Dixon, R. and Spalding, L.		
JOURNAL	The sequence of Homo sapiens BAC clone RP11-321C18		
REFERENCE	unpublished (2001)		
AUTHORS	3 (bases 1 to 185947)		
TITLE	Waterston, R.H.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (19-JUL-2001) Genome Sequencing Center, Washington		
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
TITLE	MO 63108, USA		
JOURNAL	4 (bases 1 to 185947)		
REFERENCE	Waterston, R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (05-FEB-2002) Genome Sequencing Center, Washington		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
REFERENCE	MO 63108, USA		
AUTHORS	5 (bases 1 to 185947)		
TITLE	Waterston, R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (08-FEB-2002) Department of Genetics, Washington		
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
TITLE	On Feb 5, 2002 this sequence version replaced gi:15624956.		
JOURNAL			
COMMENT			

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanesse, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pbcpc.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-564H1, 2000 bp overlap; the clone sequenced to the right is RP11-466M21, 2000 bp overlap. Actual start of this clone is at base position 173851 of RP11-564H1; actual end is at base position 9343 of RP11-466M21.

Data from AC012366 was used to finish this clone, AC092650. A single subclone region exists between 47721 and 47759.

The sequence of AC032022 has been incorporated into AC092650.

FEATURES

```
source
    Location/Qualifiers
        1..185947
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2"
            /clone="RP11-321C18"
            /clone_id="RPCI-11"
            1..318
            /note="similar to EST AW889603 (NTD:98053808)"
        misc_feature
            repeat_region
                367..577
                /rpt_family="Alu"
            repeat_region
                723..1225
                /rpt_family="MALR"
            repeat_region
                1332..1397
                /rpt_family="L2"
            repeat_region
                1398..1699
                /rpt_family="Alu"
            repeat_region
                1700..1829
                /rpt_family="L2"
            repeat_region
                2028..2067
                /rpt_family="AT-rich"
            repeat_region
                2364..2487
                /rpt_family="L1"
            repeat_region
                2483..2720
                /rpt_family="L1"
            repeat_region
                2714..2734
                /rpt_family="Alu"
            repeat_region
                2738..3021
                /rpt_family="AT-rich"
            repeat_region
                2886..2913
                /rpt_family="(T)n"
            repeat_region
                2914..3464
                /note="similar to EST A180806 (NTD:95394554) wf92c04.x1"
            repeat_region
                2914..3209
                /note="similar to EST A1470488 (NTD:94332578) tj44e05.x1"
            repeat_region
                3003..3243
                /note="similar to EST BF874556 (NTD:912264686)"
            repeat_region
                3009..3848
                /note="similar to EST BG620144 (NTD:913671505)"
            repeat_region
                3031..3056
                /rpt_family="AT-rich"
```

```
misc_feature
    3113..3650
    /note="similar to EST AW945629 (NTD:98123384)"
misc_feature
    3506..3775
    /note="similar to EST BE004742 (NTD:98264975)"
misc_feature
    3507..3994
    /note="similar to EST BG572978 (NTD:913580631)"
misc_feature
    3905..4229
    /note="similar to EST R81876 (NTD:9858479) yj05h12.r1"
repeat_region
    4041..4106
    /rpt_family="Alu"
repeat_region
    5466..5998
    /rpt_family="MER2-type"
repeat_region
    6244..6811
    /rpt_family="MALR"
misc_feature
    7420..7879
    /note="similar to EST A1921626 (NTD:95657590) wo27h06.x1"
repeat_region
    7421..7628
    /rpt_family="L1"
repeat_region
    7834..8029
    /rpt_family="MALR"
repeat_region
    8975..9220
    /rpt_family="Alu"
repeat_region
    9416..9487
    /rpt_family="(TA)n"
repeat_region
    9976..10321
    /rpt_family="MALR"
misc_feature
    10344..10772
    /note="similar to EST BF511583 (NTD:911594881)"
repeat_region
    10350..10385
    /rpt_family="L2"
repeat_region
    10641..10757
    /rpt_family="MIR"
misc_feature
    12091..12296
    /note="similar to EST AV748457 (NTD:910906305)"
misc_feature
    12149..12296
    /note="similar to EST BF993178 (NTD:912299501)"
misc_feature
    12171..12296
    /note="similar to EST BG232143 (NTD:912727307)"
repeat_region
    12565..12864
    /rpt_family="Alu"
repeat_region
    12982..13123
    /rpt_family="L1"
repeat_region
    13266..13287
    /rpt_family="AT-rich"
repeat_region
    13441..13683
    /rpt_family="L1"
repeat_region
    13684..14454
    /rpt_family="L1"
repeat_region
    14177..14263
    /rpt_family="A-rich"
repeat_region
    14455..14767
    /rpt_family="Alu"
repeat_region
    14740..14762
    /rpt_family="AT-rich"
repeat_region
    14768..14869
    /rpt_family="L1"
repeat_region
    15036..15327
    /rpt_family="Alu"
repeat_region
    15354..15390
    /rpt_family="(A)n"
repeat_region
    15795..15985
    /rpt_family="L1"
repeat_region
    16447..16470
    /rpt_family="(TTTG)n"
repeat_region
    17751..19325
    /rpt_family="L1"
repeat_region
    19329..19387
    /rpt_family="L1"
repeat_region
    19388..19679
    /rpt_family="Alu"
repeat_region
    19680..20376
    /rpt_family="L1"
```


KEYWORDS	hug.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 161221)
TITLE	Smith,D.R.
JOURNAL	Genome Therapeutics Corporation Sequencing Center: Human Genome
REFERENCE	Sequence Data
AUTHORS	Unpublished
TITLE	2 (bases 1 to 161221)
JOURNAL	Smith,D.R.
REFERENCE	Direct Submission
AUTHORS	Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
TITLE	Street, Waltham, MA 02453, USA
JOURNAL	3 (bases 1 to 161221)
REFERENCE	Smith,D.R.
AUTHORS	Direct Submission
TITLE	Submitted (21-AUG-2001) Genome Therapeutics Corporation, 100 Beaver
JOURNAL	Street, Waltham, MA 02453, USA
COMMENT	On Aug 21, 2001 this sequence version replaced gi:15011655.
FEATURES	location/Qualifiers
SOURCE	1..161221
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="10"
	/clone="RP11-131N11"
	/clone_1kb="RC1-11"
BASE COUNT	51066 a 31550 c 30263 g 48342 t
ORIGIN	

Query Match	6.8%	Score 37.6;	DB 9;	Length 161221;
Best Local Similarity	61.0%;	Pred. No. 15;		
Matches 61;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0;
QY 146	tgagacaaatcttatgacaaacggcctgctgctggaanaaatttgatcggaaac	205		
Db 77122	TGAAATATATNGTTAAATTAATTTATGCTCTTATATATGAAACACCATTAATATATAAAAAAT	77181		
QY 206	taaaagcacaacaggaactgctgctgctgagcgtaccta	245		
Db 77182	TAAAAAGAGCAACCGAATATTTTGGAGATTCGCTGTAACTTA	77221		

Search completed: June 27, 2002, 20:16:15
Job time: 9703 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 18:35:02 ; Search time 362.26 Seconds
(Without alignments)
2630.397 Million cell updates/sec

Title: US-09-727-769a-5
Perfect score: 555
Sequence: 1 ttggcgagtgtaattcctga.....atgtatccagctgtgattt 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.Geneseq_032802.*
2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	555	100.0	555	22	AAF90280
2	555	100.0	1080	22	AAF90281
3	344.2	62.0	555	21	AAZ49494
4	344.2	62.0	1080	21	AAZ49495
5	37	6.5	15865	22	AAK90825
6	36	6.5	19576	24	AAK61258
7	35.6	6.4	2489	21	AAZ90461
8	35.4	6.4	9300	22	ABA15594
9	35	6.3	4383	23	ABL18719

10	35	6.3	4590	22	AAH24065
11	35	6.3	106416	23	ABL18718
12	34.8	6.3	1751	19	AAV40300
13	34.4	6.2	1072	18	AAV74746
14	34.4	6.2	1570	22	ABA18492
15	34.4	6.2	3943	22	AAK84357
16	34.4	6.2	12582	22	AAK84357
17	34.4	6.2	23821	21	AAK92469
18	34.2	6.2	2019	22	AAK25737
19	34.2	6.2	4197	16	AAQ99430
20	34	6.1	3861	22	AAQ99377
21	34	6.1	3945	22	AAK31428
22	33.8	6.1	9289	22	AAK46502
23	33.6	6.1	4223	22	AAH78011
24	33.4	6.0	932	22	AAH29847
25	33.2	6.0	397	14	AAQ60606
26	33	5.9	3442	21	AAK59824
27	33	5.9	5195	24	ABL33596
28	32.8	5.9	1949	11	AAQ60618
29	32.8	5.9	1950	10	AAK91094
30	32.8	5.9	1950	11	AAQ06017
31	32.8	5.9	1950	12	AAQ10001
32	32.8	5.9	1950	15	AAQ64644
33	32.8	5.9	5366	20	AAK58001
34	32.8	5.9	5366	20	AAK56273
35	32.8	5.9	5502	17	AAK30092
36	32.8	5.9	5502	18	AAK71263
37	32.8	5.9	5984	20	AAK58000
38	32.8	5.9	5984	20	AAK56272
39	32.8	5.9	5984	21	AAK39808
40	32.8	5.9	6124	18	AAK71265
41	32.8	5.9	6124	18	AAK71264
42	32.8	5.9	6228	18	AAK71266
43	32.8	5.9	28473	19	AAK52216
44	32.6	5.9	13431	22	AAK08701
45	32.4	5.8	575	22	AAK30852

ALIGNMENTS

RESULT	ID	Score	Description
1	AAF90280	555	AAF90280 standard; DNA: 555 BP.
XX	AAF90280:		
XX	22-AUG-2001 (first entry)		
DT			
XX			
DE			Nucleotide sequence of a protein-deamidating enzyme.
XX			
KW			Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KX			bakery; confectionery; ss.
OS			Cryoseobacterium sp.
XX			
PN			EP1106696-A1.
XX			
PD			13-JUN-2001.
XX			
PF			04-DEC-2000; 2000EP-0310768.
XX			
PR			03-DEC-1999; 99JP-0345044.
XX			
PA			(AMAN-) AMANO ENZYME INC.
XX			
PI			Yamaguchi S;
XX			
DR			WPI: 2001-376907/40.
XX			P-P-SDB; AAB64386.
PT			New enzyme for use in e.g. bakery has an ability to deamidate amido groups in a protein

Yeast AOD9604-asso
Drosophila melanog
Sugarbeet choline
Staphylococcus aur
Human nervous syst
Human cDNA encodin
Human immune/haema
Shewanella sp. SCR
Tomato LIR7 promot
B. sphaericus SLP
Human vesicle traf
Human cDNA encodin
Tumour suppressor
Nucleotide sequenc
S cerevisiae apopt
Human brain Expres
Human secreted pro
Human immune syste
Sequence encoding
Protein G gene. S
Sequence encoding
Streptococcus Prot
Gonadotropic hormo
Human apoptosis in
Neutrol apoptosis
Neutrol apoptosis
Gonadotropic hormo
Human apoptosis in
Human NAIP cDNA.
Neutrol apoptosis
Neutrol apoptosis
Neutrol apoptosis
Streptococcus pneu
Lycopodium pennu
Human cDNA encodin

XX Claim 8; Page 22; 43pp; English.
XX
PS The present sequence encodes a protein-deamidating enzyme from
CC Cryoseobacterium sp. number 9670. The enzyme is able to deaminate amido
CC groups in a protein by directly acting upon the amido groups without
CC cutting peptide bonds and without cross-linking the protein. The enzyme
CC thus reduces the mineral sensitivity of the protein and increases the
CC soluble mineral content in the protein-mineral solution, improving the
CC absorption of minerals in the human body. The enzyme reduces or removes
CC the toxicity and allergenic property of the protein in a food
CC e.g. allergy. The enzyme is useful for the improvement of dough in the
CC field of bakery and confectionery e.g. for the production of crackers,
CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC protein products; in various food articles e.g. meat or fish products
CC and noodles; and for improving functionality of plant or animal protein.
XX
SQ Sequence 555 BP; 166 A; 119 C; 107 G; 163 T; 0 other;

Query Match 100.0%; Score 555; DB 22; Length 555;
Best Local Similarity 100.0%; Pred. No. 4.2e-160;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttggcgagtgtaattccctgagtagctacataaattcttattcaatcaataagaat 60
DB 1 ttggcgagtgtaattccctgagtagctacataaattcttattcaatcaataagaat 60
OY 61 cagcttcggtgtaacctgaagcgctccacatgacatgacatgacatgacatgac 120
DB 61 cagcttcggtgtaacctgaagcgctccacatgacatgacatgacatgacatgac 120
OY 121 ggaagttaatgcaagagccataagaatgagacaatcttaatgaacaagcgctatgact 180
DB 121 ggaagttaatgcaagagccataagaatgagacaatcttaatgaacaagcgctatgact 180
OY 181 gaaacaatttgatgaaggaacctaaagatcaacaggaactgctgtgtgctgtg 240
DB 181 gaaacaatttgatgaaggaacctaaagatcaacaggaactgctgtgtgctgtg 240
OY 241 agctacccggtgacataatgtaagcataaaatgcttcggagtaacggaaaaaaga 300
DB 241 agctacccggtgacataatgtaagcataaaatgcttcggagtaacggaaaaaaga 300
OY 301 attattgaccccttcaatttcaagaagcgctccgttaacagabacagcatgagaacgct 360
DB 301 attattgaccccttcaatttcaagaagcgctccgttaacagabacagcatgagaacgct 360
OY 361 tgcgttaacacctcttgagatctgacatccgtttctcttattgacataactgcaagaat 420
DB 361 tgcgttaacacctcttgagatctgacatccgtttctcttattgacataactgcaagaat 420
OY 421 gttatttcagaagtcacagtaattcttaccctgatagaacaactctgatactccaac 480
DB 421 gttatttcagaagtcacagtaattcttaccctgatagaacaactctgatactccaac 480
OY 481 tgtgtactgactaaatttcaactgcttcggagtgcttctccttaccctgacaggaatgta 540
DB 481 tgtgtactgactaaatttcaactgcttcggagtgcttctccttaccctgacaggaatgta 540
OY 541 tccagctgtgattt 555
DB 541 tccagctgtgattt 555

RESULT 2

AAF90281

ID AAF90281 standard; DNA: 1080 BP.

XX AAF90281;
XX 22-AUG-2001 (first entry)
XX

DE Nucleotide sequence of a protein-deamidating enzyme.
XX
XX Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery; ss.
XX
OS Cryoseobacterium sp.
XX
FH Key Location/Qualifiers
FT CDS 61..1023
FT /tag="a
FT /product="protein-deamidating enzyme"
XX
XX EPI106696-AI.
XX
XX 13-JUN-2001.
XX
XX 04-DEC-2000; 2000EP-0310768.
XX
XX 03-DEC-1999; 99JP-0345044.
XX
XX (AMANO) AMANO ENZYME INC.
XX
XX Yamaguchi S;
XX
XX WPI: 2001-376907/40.
XX
XX P-PSDB: AAB84387.
XX
XX New enzyme for use in e.g. bakery has an ability to deaminate amido
XX groups in a protein -
XX
XX Example 11; Page 23; 43pp; English.
XX
XX The present sequence encodes a protein-deamidating enzyme from
CC Cryoseobacterium sp. number 9670. The enzyme is able to deaminate amido
CC groups in a protein by directly acting upon the amido groups without
CC cutting peptide bonds and without cross-linking the protein. The enzyme
CC thus reduces the mineral sensitivity of the protein and increases the
CC soluble mineral content in the protein-mineral solution, improving the
CC absorption of minerals in the human body. The enzyme reduces or removes
CC the toxicity and allergenic property of the protein in a food
CC e.g. allergy. The enzyme is useful for the improvement of dough in the
CC field of bakery and confectionery e.g. for the production of crackers,
CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean
CC protein products; in various food articles e.g. meat or fish products
CC and noodles; and for improving functionality of plant or animal protein.
XX
XX
SQ Sequence 1080 BP; 371 A; 204 C; 203 G; 302 T; 0 other;

Query Match 100.0%; Score 555; DB 22; Length 1080;
Best Local Similarity 100.0%; Pred. No. 5.6e-160;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttggcgagtgtaattccctgagtagctacataaattcttattcaatcaataagaat 60
DB 466 ttggcgagtgtaattccctgagtagctacataaattcttattcaatcaataagaat 525
OY 61 cagcttcggtgtaacctgaagcgctccacatgacatgacatgacatgacatgac 120
DB 526 cagcttcggtgtaacctgaagcgctccacatgacatgacatgacatgacatgac 585
OY 121 ggaagttaatgcaagagccataagaatgagacaatcttaatgaacaagcgctatgact 180
DB 586 ggaagttaatgcaagagccataagaatgagacaatcttaatgaacaagcgctatgact 645
OY 181 gaaacaatttgatgaaggaacctaaagatcaacaggaactgctgtgtgctgtg 240
DB 646 gaaacaatttgatgaaggaacctaaagatcaacaggaactgctgtgtgctgtg 705
OY 241 agctacccggtgacataatgtaagcataaaatgcttcggagtaacggaaaaaaga 300
DB 706 agctacccggtgacataatgtaagcataaaatgcttcggagtaacggaaaaaaga 765

QY 301 attattgatccctcactattttcaagcgcgtcgttaacagatagagaaagcgt 360
|||||
Db 766 attattgatccctcactattttcaagcgcgtcgttaacagatagagaaagcgt 825
QY 361 tgcgttaaacctcttgcgatcgtcgtcgttccctctatgtaatactgaggaat 420
|||||
Db 826 tgcgttaaacctcttgcgatcgtcgtcgttccctctatgtaatactgaggaat 885
QY 421 gttattacagaagctcctaattcttaccgtatagacaacaatcgtatcaatacgaac 480
|||||
Db 886 gttattacagaagctcctaattcttaccgtatagacaacaatcgtatcaatacgaac 945
QY 481 tgttactgactaaatttctactgcttccgagatgtctctcttaccctgacccgagatgta 540
Db 946 tgttactgactaaatttctactgcttccgagatgtctctcttaccctgacccgagatgta 1005
QY 541 tccagctgtgattt 555
|||||
Db 1006 tccagctgtgattt 1020

RESULT 3

AAZ49494
ID AAZ49494 standard; DNA: 555 BP.

XX AAZ49494;

DT 04-APR-2000 (first entry)

XX Chryseobacterium gleum protein-deamidating enzyme encoding DNA.

DE Protein deamidating enzyme; soil bacterium; deamidation activity;
KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
KM protein engineering; surface hydrophobicity; toxicity; allergic;
KM mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.

OS Chryseobacterium gleum 'JCM 2410'.

XX Key Location/Qualifiers

FH mat_peptide 1..555

FT /*tag- a /product- "Protein deamidating enzyme"

FT /function- "Deamidate amido groups in a protein"

FT /note= "Improves protein function"

XX EP976829-A2.

XX 02-FEB-2000.

XX 04-JUN-1999; 99EP-0304367.

XX 04-JUN-1998; 98JP-0173940.

PA (AMAN) AMANO PHARM KK.

PI Yamaguchi S, Matsuura A;

XX WPI; 2000-118552/11.

DR P-PDB; AAY44582.

XX New enzyme for modifying and improving the function of proteins and/or
PT peptides has deamidating activity without causing cross linking -

XX Claim 8; Page 23; 57pp; English.

XX The present sequence is the DNA encoding the protein-deamidating enzyme,
CC isolated from a new strain of soil bacterium, Chryseobacterium gleum
CC JCM 2410. The enzyme exerts the deamidation activity by directly acting
CC upon side chain amido groups in the protein in bonded state and
CC releasing side chain carboxyl groups and ammonia. It can deaminate high
CC molecular weight proteins, without cross linking and cleavage of peptide
CC bonds, to improve protein function. This sequence is used for protein
CC engineering, to cause an increase in surface hydrophobicity and improve

CC the function of a plant or animal protein. It can also be used to remove
CC or reduce toxicity or allergic property of proteins in food, decrease
CC mineral sensitivity of protein, to allow greater absorption into the body
CC and to solubilise calcium for use in drinks and mineral enhancing agents.
XX Sequence 555 BP; 159 A; 122 C; 112 G; 162 T; 0 other;

Query Match 62.0%; Score 344.2; DB 21; Length 555;
Best Local Similarity 76.7%; Pred. No. 1.9e-95;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 7 agtgaattcctgtagttagtaccataatcttataccaataaagaatcaatct 66
|||||
Db 7 agtgaattcctgtagttagtaccataatcttataccaataaagaatcaatct 66
QY 67 tgcgttacctcctacagcgtctccacatgcatcaatcagatatacctgtagacgagt 126
|||||
Db 67 tgcgttacctcctacagcgtctccacatgcatcaatcagatatacctgtgtacgagt 126
QY 127 tatgcaagagcccaataagatgagacaatcttaatgacaacagcgtatgtagaaaaa 186
|||||
Db 127 tatgcaagagcccaataagatgagacaatcttaatgacaacagcgtatgtagaaaaa 186
QY 187 caattgtatacaggaacaaacataaagcacaacaggaactgtgtgtgagctgac 246
|||||
Db 187 caattgtatacaggaacaaacataaagcacaacaggaactgtgtgtgagctgac 246
QY 247 caagtgaatatgtgtaagctataaaatgcttcgagatgaacggaataaagaattact 306
|||||
Db 247 caagtgaatatgtgtaagctataaaatgcttcgagatgaacggaataaagaattact 306
QY 307 gatcttcactatttttaagcagcgtcgttaacagatatacagatggaataacgcttgct 366
|||||
Db 307 gatcttcactatttttaagcagcgtcgttaacagatatacagatggaataacgcttgct 366
QY 367 aacaccttcgagatcgtacatccgttccctctatgctaatactgacaggaatgttat 426
|||||
Db 367 aacaccttcgagatcgtacatccgttccctctatgctaatactgacaggaatgttat 426
QY 427 tacagaagctcctaataatcttaccgtatgacaacaatcgtatcaatcacaactgtgta 486
|||||
Db 427 tacagaagctcctaataatcttaccgtatgacaacaatcgtatcaatcacaactgtgta 486
QY 487 ctgactaattttacagcttccgagatgtctctcttaccctgacccgagatgatacagc 546
|||||
Db 487 ctgactaattttacagcttccgagatgtctctcttaccctgacccgagatgatacagc 546
QY 547 tgtgattt 555
|||||
Db 547 tgtgattt 555

RESULT 4

AAZ49495
ID AAZ49495 standard; DNA: 1080 BP.

XX AAZ49495;

DT 04-APR-2000 (first entry)

XX Chryseobacterium gleum protein-deamidating prepro-enzyme encoding gene.

DE Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;

KW amido group; carboxyl group; ammonia; cross linking; peptide bond;

KM protein engineering; surface hydrophobicity; toxicity; allergic;

KM mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.

OS Chryseobacterium gleum 'JCM 2410'.

XX Key Location/Qualifiers

FH CDS 61..1020

FT /*tag- a

```
FT      /product- "C. gleum protein deamidating prepro-enzyme"
FT      /function- "deamidate amido groups in a protein"
FT      sig_peptide
FT      61..123
FT      /*tag- b
FT      /note- "Corresponds to pre region of the protein"
FT      mat_peptide
FT      463..1017
FT      /*tag- C
FT      /label- Mature_protein_deamidating_enzyme
FT      /note- "Improves protein function"
FT
FT      EP976829-A2.
FT
FT      02-FEB-2000.
FT
FT      04-JUN-1999; 99EP-0304367.
FT
FT      04-JUN-1998; 98JP-0173940.
FT
FT      (AMANO ) AMANO PHARM KK.
FT
FT      Yamaguchi S, Matsuura A;
FT      WPI: 2000-118552/11.
FT      P-PSDB: AAY44583.
FT
FT      New enzyme for modifying and improving the function of proteins and/or
FT      peptides has deamidating activity without causing cross linking -
FT
FT      Example 26; Page 25; 57pp; English.
FT
FT      The present sequence is the gene encoding the protein-deamidating
FT      prepro-enzyme, isolated from a new strain of soil bacterium,
FT      Chryseobacterium gleum JCM 2410. The enzyme exerts the deamidation
FT      activity by directly acting upon side chain amido groups in the protein
FT      in bonded state and releasing side chain carboxyl groups and ammonia. It
FT      can deaminate high molecular weight proteins, without cross linking and
FT      cleavage of peptide bonds, to improve protein function. This sequence is
FT      used for protein engineering, to cause an increase in surface
FT      hydrophobicity and improve the function of a plant or animal protein. It
FT      can also be used to remove or reduce toxicity or allergic property of
FT      proteins in food, decrease mineral sensitivity of protein, to allow
FT      greater absorption into the body and to solubilise calcium for use in
FT      drinks and mineral enhancing agents.
FT
FT      Sequence 1080 BP; 354 A; 210 C; 205 G; 311 T; 0 other;
FT
FT      Query Match      62.0%; Score 344.2; DB 21; Length 1080;
FT      Best Local Similarity 76.7%; Pred. No. 2.5e-95;
FT      Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
```

```
Db      769 gattccttaattctccagcgctctctgtaacaagattctgcatgagagctgcatgtacc 828
Qy      367 aacctcttcgagatcgcgtacccgttccctcttactgtaataacgcaggaattttat 426
Db      829 aacacaagctcgagatcgcgtctgcttactcttccacgaacacgcaggaattttac 888
Qy      427 tacagaagctcgtagtaattcttaacctgtatgacaacaatctgataataccaactgtgta 486
Db      889 tacagaagctcgtacgattcttaactgtatgataacaactatgtaataccaattgtgta 948
Qy      487 ctgactaaatttcactctcttcgagatcttcccttaccacgcagcgatgtaccagc 546
Db      949 ttaacacatctctccatcccttccagagatctctccctccccaagcaagtgtaagcaagc 1008
Qy      547 tctgattt 555
Db      1009 tctgattt 1017

RESULT 5
ID      AAK90825/C
ID      AAK90825 standard; DNA; 15865 BP.
AC      AAK90825;
AC
DE      05-NOV-2001 (first entry)
DE
DE      Human digestive system antigen genomic sequence SPQ ID NO: 4401.
KW      Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW      ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW      digestive system disorder; Meckel's diverticulum; ds.
OS      Homo sapiens.
XX      WO200155314-A2.
XX
XX      02-AUG-2001.
XX
XX      17-JAN-2001; 2001WO-US01324.
XX
PR      31-JAN-2000; 2000US-0179065.
PR      04-FEB-2000; 2000US-0180628.
PR      24-FEB-2000; 2000US-0184664.
PR      02-MAR-2000; 2000US-0186350.
PR      16-MAR-2000; 2000US-0189874.
PR      17-MAR-2000; 2000US-0190076.
PR      18-APR-2000; 2000US-0198123.
PR      19-MAY-2000; 2000US-0205515.
PR      07-JUN-2000; 2000US-0209467.
PR      28-JUN-2000; 2000US-0214886.
PR      30-JUN-2000; 2000US-0215135.
PR      07-JUL-2000; 2000US-0216647.
PR      07-JUL-2000; 2000US-0216880.
PR      11-JUL-2000; 2000US-0217487.
PR      11-JUL-2000; 2000US-0217496.
PR      14-JUL-2000; 2000US-0218290.
PR      26-JUL-2000; 2000US-0220963.
PR      26-JUL-2000; 2000US-0220964.
PR      14-AUG-2000; 2000US-0224518.
PR      14-AUG-2000; 2000US-0224519.
PR      14-AUG-2000; 2000US-0225213.
PR      14-AUG-2000; 2000US-0225214.
PR      14-AUG-2000; 2000US-0225266.
PR      14-AUG-2000; 2000US-0225267.
PR      14-AUG-2000; 2000US-0225268.
PR      14-AUG-2000; 2000US-0225270.
PR      14-AUG-2000; 2000US-0225447.
PR      14-AUG-2000; 2000US-0225757.
PR      14-AUG-2000; 2000US-0225758.
PR      14-AUG-2000; 2000US-0225759.
PR      18-AUG-2000; 2000US-0226279.
PR      22-AUG-2000; 2000US-0226681.
```

PR	22-AUG-2000	2000US-0226668
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227709
PR	30-AGO-2000	2000US-0228924
PR	01-SEP-2000	2000US-0228987
PR	01-SEP-2000	2000US-0229143
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229519
PR	05-SEP-2000	2000US-0229530
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231143
PR	08-SEP-2000	2000US-0231144
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	14-SEP-2000	2000US-0232939
PR	14-SEP-2000	2000US-0232940
PR	14-SEP-2000	2000US-0232941
PR	14-SEP-2000	2000US-0232942
PR	14-SEP-2000	2000US-0232943
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	21-SEP-2000	2000US-0234397
PR	21-SEP-2000	2000US-0234398
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236519
PR	02-OCT-2000	2000US-0236670
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239353
PR	13-OCT-2000	2000US-0239357
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241865
PR	20-OCT-2000	2000US-0241867
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241816
PR	01-NOV-2000	2000US-0244523
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	17-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207

[illegible]

AA61258/c	standard; DNA; 19576 BP.
XX	AA61258:
AC	29-JAN-2002 (first entry)
XX	
DT	Human gene regulation-associated gene oligonucleotide #213.
DE	
XX	
KW	Human: Gene regulation-associated gene: severe combined immunodeficiency
KW	cardiac damage; inflammatory response; Hemophilia; Werner syndrome;
KW	asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW	renal disease; Preeclampsia; cardiac allograft vascular disease;
KW	colorectal cancer; thyroid cancer; oesophageal cancer; ds: tumour;
KW	immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;
KW	nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX	
OS	Homo sapiens.
XX	
PN	MO200177375-A2.
XX	
PD	18-OCT-2001.
XX	
PF	06-APR-2001; 2001MO-EP03968.
XX	
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIC-) EPIDEMIOICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
DR	WPI: 2002-017470/02.
XX	
PT	New nucleic acid sequences from chemically modified genes associated
PT	with gene regulation, useful for analysing cytosine methylations for
PT	diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT	disease
XX	
PS	Claim 1; SEQ ID NO 219; 26pp; English.
XX	
CC	The invention relates to 224 nucleic acid sequences comprising at least
CC	18 bases of a chemically pretreated gene associated with gene regulation
CC	selected from 43 known genes (or complementary sequences). The
CC	chemical pretreatment converts cytosine bases unmethylated at the
CC	5-position to uracil or another base with hybridisation behaviour
CC	dissimilar to cytosine, to enable analysis of cytosine methylations.
CC	The DNA sequences, oligomers (or sets/arrays) and method are
CC	useful in the diagnosis of diseases (or predisposition to diseases)
CC	associated with gene regulation and in therapy of such diseases, by
CC	enabling analysis of the cytosine methylation patterns of such genes,
CC	kits are provided. They are especially useful in diagnosis
CC	and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC	disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC	asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC	preeclampsia, graft versus-host disease. The present sequence is a
CC	sequence included in the sequence data for this specification and is
CC	associated with the human gene regulation-associated genes.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pat_sequences
XX	
90	Sequence 19576 BP; 5357 A; 307 C; 4161 G; 9750 T; 1 other;

Query Match	6.5%	Score 36;	DB 24;	Length 19576;
Best Local Similarity	49.5%;	Pred. NO. 3;		
Matches 93;	Conservative 0;	Mismatches 95;	Indels 0;	Gaps 0;
Oy	312	ttcactatttcaagcgtctctgttaacagatcacagatgagaaacgcttgcyttaaac	371	

Db	6024	TTCTCTATTATTTAAATTAATCTACCTACCTATTATTAATTAACATTCATTAATTCTACCT	5965
Qy	372	ctctgcggagatcgcacatccgtttctctcttaatgctaactgcaggaaatgttattacag	431
Db	5964	AACGTAAAAACATTTAATCTTTTGGCTGTAAAAATATATATATATATATTTATTAATAA	5905
Qy	432	aagctctagtatcttactcctcgtctgtgacacaaatcgtatcaataccaactgltactgac	491
Db	5904	AAAAATAATCACTCTTTTAACCAAAATATTAATAATTTCTAATTAATAATCTACTATCCAC	5845
Oy	492	taaatctt 499	
Db	5844	AAAAATTT 5837	
RESULT	7		
AA290461/C			
ID	AA290461	standard; DNA; 2489 BP.	
AC	AA290461;		
XX			
DT	06-JUN-2000	(first entry)	
XX			
DE	E. canis 120 kDa immunodominant antigenic protein encoding DNA.		
XX			
KW	120 kDa protein; immunodominant; antigen; immunoreactive; vaccine;		
KW	Ehrlichia canis infection; antibacterial1; ds.		
XX			
OS	Ehrlichia canis.		
XX			
EH	Key	Location/Qualifiers	
FT	CDS	341..2407	
FT		/*tag= a	
FT		/product= "120 kDa immunodominant antigenic protein"	
PN	WO200012688-A1.		
XX			
PD	09-MAR-2000.		
XX			
PF	27-AUG-1999;	99WO-US19538.	
XX			
PR	27-AUG-1998;	98US-0141047.	
XX			
PA	(RERE-) RES DEV FOUND.		
XX			
PI	Walker DH, Yu X;		
XX			
DR	WPI: 2000-256636/22.		
DR	P-PSDB; AAY57274.		
PT	Protein immunoreactive with anti-Ehrlichia canis and comprises a		
PT	sequence of 688 amino acids, useful for inhibiting Ehrlichia canis		
PT	infection		
XX			
PS	Claim 2; Fig 3; 78pp; English.		
XX			
CC	The invention provides a 120 kDa immunodominant antigenic protein of		
CC	Ehrlichia canis that is immunoreactive with anti-Ehrlichia canis		
CC	serum. The protein can be expressed by standard recombinant methodology		
CC	The protein is useful for inhibiting Ehrlichia canis infection. The		
CC	present sequence represents a DNA encoding the immunodominant protein		
CC	of E. canis.		
XX			
QO	Sequence 2489 BP; 935 A; 286 C; 552 G; 716 T; 0 other;		

	Query Match	6.4%	Score 35.6	DB 21	Length 2489
	Best Local Similarity	45.9%	Pred. NO. 1.6		
	Matches 122	Conservative 0	Mismatches 144	Indels 0	Gaps 0
QY	276	tgcttcgagagtaacgcgaaaaagaattatgtatcccttcaactatttcaagcgcttcgt	335		

PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250392.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251999.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-541565/60.
XX DR Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases -
XX PS Disclosure: SEQ ID NO 7925; 1701pp + Sequence Listing; English.
XX CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX SQ Sequence 9300 BP; 2858 A; 1566 C; 1864 G; 3012 T; 0 other:

Query Match 6.4%; Score 35.4; DB 22; Length 9300;
Best Local Similarity 49.2%; Pred. No. 3.3; Mismatches 96; Indels 0; Gaps 0;
Matches 93; Conservative 0;

Oy 7 agtgaattccgtagctacataaattcttattcaatcaataaagaatcagctc 66
Db 5954 AATTTTTTTTTTTTNGTGAGAGCAAAAGGCTTGATGTACCCAAAGCTGCTGATCT 5895
Oy 67 tgggtacctctacggcgctccaccatgcatcacatcagatctcctgtagaggaatc 126
Db 5894 CCTGGTGCCAAACAAACCTCCACCCTGGCCCTCCATTTTAAAAAATTTTGTGATCTACT 5835
Oy 127 tatgaagagcccaataagatgagacaatcttaagaacaagctatgacgtgtaaaa 186
Db 5834 CATTCAAGGGCCTTTTAAATTAAGAACATTTACATTTTCATTTTCTGTCAATTTTAAAAA 5775
Oy 187 caattgtca 195
Db 5774 AATTTTAA 5766
RESULT 9
ABLI8719
ID ABLI8719 standard; DNA; 4383 BP.
XX ABLI8719;
AC ABLI8719;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 7630.
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7630.
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2001; 2001WO-US09231.
PF 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX DR WPI: 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX PS Claim 1; SEQ ID NO 7630; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBI57737-ABBI2072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX SQ Sequence 4383 BP; 1563 A; 942 C; 865 G; 1013 T; 0 other:

Query Match 6.3%; Score 35; DB 23; Length 4383;
Best Local Similarity 57.9%; Pred. No. 3.2; Mismatches 45; Indels 0; Gaps 0;
Matches 62; Conservative 0;

```
OY 93 atgcatacattcagatctccttagacgagatgtatgcaagaccataagatgagaca 152
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 654 atgaaccacagtaacatgattgttcagcaactactactaatacacgagtataccaagaa 713
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 153 aatcttatgacaacgagctagtctgtgaacaacatttgtatcacg 199
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 714 catcaacaacacacactgcaacgcaaccaacaacatttgtatcacg 760

RESULT 10
AAH24065
ID AAH24065 standard; DNA; 4590 BP.
AC AAH24065;
XX
XX 29-AUG-2001 (first entry)
DE Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
XX
XX Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
KW modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
KW lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
KW functional food; transgenic yeast; fat/lean ratio; food use; ds.
XX
XX Saccharomyces cerevisiae.
OS
XX
XX Key Location/Qualifiers
FH 10
FT misc_feature /tag- a "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3617 /tag- b "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3649 /tag- c "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3679 /tag- d "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3819 /tag- e "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3862 /tag- f "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3864 /tag- g "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3888 /tag- h "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3890 /tag- i "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3912 /tag- j "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3914 /tag- k "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3938 /tag- l "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3939 /tag- m "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3941 /tag- o "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 3943 /tag- p "Represented as * in the specification"
FT /note- "Represented as * in the specification"
FT 4361 misc_feature
```

```
FT /tag- g
FT /note- "Represented as * in the specification"
PN W0200133977-A1.
PD 17-MAY-2001.
PF 06-NOV-2000; 2000MO-AU01362.
XX
XX 05-NOV-1999; 99AU-0003875.
XX (META-) METABOLIC PHARM LTD.
XX Belyea CI, Ng FM, Vaughan P;
XX WPI; 2001-328876/34.
XX
XX New organisms containing nucleic acid encoding a growth hormone
PT fragment which modulates lipid metabolism are useful to produce dietary
PT aids for obesity and in the meat production industry
XX
XX Disclosure; Page 48-50; 54pp; English.
XX
XX The invention relates to novel transgenic organisms useful in the
CC production of functional food and drink products for the treatment
CC or prevention of obesity via the regulation of lipid metabolism. The
CC organisms comprise a polynucleotide encoding a growth hormone fragment
CC capable of stimulating the activity of hormone-sensitive lipase (the key
CC enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
CC at least the disulphide-bonded loop of a mammalian growth hormone (but is
CC not the full-length growth hormone) and is optionally linked to an
CC epitope tag or heterologous fusion protein partner. The transgenic
CC organism may be a microorganism used to produce a fermented product
CC (e.g., yeast), or an edible plant or animal or cell thereof. Food or
CC drink made using methods of the invention are used to modify fat/lean
CC ratio, lipid metabolism or food use in a mammal. In particular, the food
CC or drink products may be used to treat or prevent obesity, particularly
CC in humans, and may also be used to improve the fat/lean ration of
CC livestock raised for meat production. In the exemplification of the
CC invention, the human growth hormone (hGH) fragment analogue AOD9604 was
CC expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
CC The present sequence is described as a DNA sequence from yeast in
CC the sequence listing, but is not further referred to in the
CC specification.
XX
SQ Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;

Query Match 6.3%; Score 35; DB 22; Length 4590;
Best Local Similarity 11.2%; Pred. No. 3.2;
Matches 53; Conservative 226; Mismatches 192; Indels 2; Gaps 1;

OY 34 aatcttatcaatcaataaagaatcagcttcggtacctcagcgctccaca 93
    . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3880 rvhdcdvbnhvvddhdhdhdhdhndwchnhbrvhdhdhddvymrgmhnnh 3939
    . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 94 tgcatacatcagatatcctgtagacgagatgataagaccataagatgagaca 153
    : : | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3940 nhntcdhdhndhntchabdqkmvvvvvvvvvvvvvchkhndhkhmchmccm 3999
    : : | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 154 atcttaatgaacagcgctagctgtgaacaacattgtatacgggaacataagca 213
    : : : : : | : : : : : | : : | : | : | : | : | : : : : : : : :
Db 4000 hkmhnaaaanmcmhnhadngckhnadataatagrahthdrtrragmbrmmbrs- 4058
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 214 tcaacaggaactgctgtgctgtagagctaccacgcttgcaatatggtataagaa 273
    | : : : : : : : : : : | : : : : : : : : : : : : : : : : : : :
Db 4059 -cdmmcmkmnmkwykndbdwryytlgbddrkwnmdsbshnmwdskdcdwvdsctmat 4117
    : : : : : : : : : : | : : : : : : : : : : : : : : : : : : :
OY 274 aatgcttcgaggaagaaagaatatattatcctccatcattcagcggtcct 333
    : : : : : : : : : : | : : : : : : : : : : : : : : : : : : :
Db 4118 rcvrysavasdmdvdsctmatrcvrysavasdmdvdsctmatrcvrysavasdmdvdsct 4177
```

```

Db 104703 CATCAACAACAACTGCAACGGCACACACAACTAATTGTATCG 1046557
||||| | ||||| | | | | ||||| |||||
RESULT 12
ID AAV40300 standard; cDNA: 1751 BP.
XX AAV40300;
XX
DT 15-OCT-1998 (first entry)
DE Sugarbeet choline monooxygenase encoding cDNA.
XX
KM Sugarbeet; choline monooxygenase; CMO; transgenic plant; crop;
KM stress tolerance; glycine betaine; pathogen; nutrition; ss.
XX
OS Beta vulgaris.
XX
FH Key Location/Qualifiers
FT CDS 37..1377
FT /*tag= a
FT /product= "choline monooxygenase"
XX
PN W09830702-A2.
XX
PD 16-JUL-1998.
XX
PF 08-JAN-1998; 98WO-US00164.
PR 08-JAN-1997; 97US-0035147.
XX
PA (UYFL ) UNIV FLORIDA.
PI Burnet M, Hanson AD, Rathinasabapathi B;
DR WPJ: 1998-399147/34.
XX P-PSDB; AAM69563.
XX
PT New polynucleotide molecule encoding choline mono:oxidase - used to,
PT e.g, transform plants to provide increased resistance to stressful
PT conditions, increase pest resistance or Increase feed value
XX
PS Claim 45; Fig 4; 36pp; English.
XX
CC The present sequence encodes sugarbeet choline monooxygenase (CMO). A
CC process has been developed for increasing the resistance of plants to
CC stressful conditions, comprising incorporating a nucleotide sequence
CC encoding CMO into a genome of the plant where the polynucleotide is
CC expressed and the plant produces glycine betaine. A method has also been
CC developed of decreasing the production of glycine betaine in a plant,
CC comprising incorporating a nucleotide sequence encoding CMO in antisense
CC form into the genome of the plant where the antisense polynucleotide is
CC expressed and the plant produces no or less glycine betaine. The
CC expression of CMO in plants can provide plants with resistance to
CC stressful conditions by osmotic adjustment and osmoprotection. The
CC blocking of the synthesis of glycine betaine can be used to improve a
CC plant's resistance to pests or pathogens. Blocking CMO in crop species
CC used as animal feed may also improve their nutritional value. The
CC blocking of glycine betaine synthesis can also facilitate industrial
CC processing, as in sugarbeet, where glycine betaine is an unwanted side
CC product interfering with sugar crystallisation. The methods can be used
CC with plants such as tobacco, wheat, barley, corn, sugarcane, sugar beet,
CC spinach, cotton, sunflower, rice, sorghum, tomato, potato, lettuce,
CC oilseed rape, or citrus.
XX
SQ Sequence 1751 BP; 547 A; 344 C; 336 G; 524 T; 0 other;
Query Match 6.3%; Score 34.8; DB 19; Length 1751;
Best Local Similarity 48.1%; Pred. No. 2.4;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

CC computer readable medium.

PR 14-AUG-2000; 2000US-0225759.

```
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234222.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241828.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251988.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX Disclosure; SEQ ID NO 10823; 1701bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABAI1004-ABAI2134) and proteins
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1570 BP; 509 A; 230 C; 234 G; 597 T; 0 other;
```

```
Query Match 5.2%; Score 34.4; DB 22; Length 1570;
Best Local Similarity 45.5%; Pred. No. 3.1;
Matches 122; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
Oy 127 tatgcaagagccataagatgagacaacitaaatgaacaacgctatgactgtgaaaaa 186
```

Db 666 TATTAACAATCACACATTTGAGGTAATTAAGTGGCAAAAAAATAATATATAT 607
QY 187 caatttgtatagcgaaccctaaagcctcaacggaaactctgctgtggcgtgagctac 246
Db 606 ATATATATACCTATATACATGACATCCAGAAATATTCACAAATCTATTTGCCCA 547
QY 247 cagcttgcaatctgtgaagctataaaatgcttcgcggagtaacggaataaattatt 306
Db 546 CACCAGCAGATAGTCAATATTAATGAAAAATATCACAATACCAACAGAAAAAGTTACACA 487
QY 307 gactcttcaactatttcaagcggctcctgttaacagatacagcatgagagaacgttcgtt 366
Db 486 TAGGGAATATACAAAAATATGTTGCTATATATCTTAAATATAGGAGTATGTTGAAT 427
QY 367 aacacctctgcggatctgcacccgtt 394
Db 426 AATGTTTGGAGAAAGTGTTCATTT 399

RESULT 15

AAS31317/c

ID AAS31317 standard; cDNA; 3943 BP.

XX AAS31317;

DT 04-DEC-2001 (first entry)

DE Human cDNA encoding a novel extracellular matrix protein, Seq ID No 131.

XX Human; secreted extracellular matrix protein; ss; immunomodulatory;
XX Anti-HIV; antineuritic; antirheumatic; antisclerotic; cardiant; vascular;
XX cerebroprotective; thrombolytic; antineurolytic; ophthalmic; cytoskeletal;
XX antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
XX human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
XX cancers; hyperproliferative disorder; breast neoplasm; melanoma;
XX Sezary syndrome; Gaucher's disease; neurological diseases;
XX Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
XX cardiac arrest; tachycardia; angina; infection; corneal infections;
XX wound healing; immunogen; gene therapy; antisense; food additive.

OS Homo sapiens.

PN WO200155368-A1.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01348.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-022547.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244517.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000: 2000US-0246527.
PR 08-NOV-2000: 2000US-0246528.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246539.
PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249219.
PR 17-NOV-2000: 2000US-0249220.
PR 17-NOV-2000: 2000US-0249221.
PR 17-NOV-2000: 2000US-0249222.
PR 17-NOV-2000: 2000US-0249223.
PR 17-NOV-2000: 2000US-0249224.
PR 17-NOV-2000: 2000US-0249225.
PR 17-NOV-2000: 2000US-0249226.
PR 17-NOV-2000: 2000US-0249227.
PR 17-NOV-2000: 2000US-0249228.
PR 17-NOV-2000: 2000US-0249229.
PR 17-NOV-2000: 2000US-0249230.
PR 17-NOV-2000: 2000US-0250160.
PR 01-DEC-2000: 2000US-0250161.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 06-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251868.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251889.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0251097.
PR 05-JAN-2001: 2001US-0251967B.

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-465572/50.
PT P-PSDB; AAU19746.
XX

Nucleic acid molecules encoding human secreted extracellular matrix proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

Claim 1: SEQ ID NO 131; 577bp; English.

The invention relates to isolated nucleic acid molecules encoding novel human secreted extracellular matrix proteins (SPs). The polynucleotides and proteins are used to prevent, treat a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For example, disorders associated with decreased expression of SPs. The SP polynucleotide or a vector expressing them may be administered to treat diseases by gene therapy. Antisense molecules may be administered to down regulate expression of SPs by binding with the cells own genes and preventing their expression. The polynucleotide may also be used as DNA probes in diagnostic assays. The SPs may also be used as antigens to produce antibodies and to identify modulators (agonists and antagonists) of the SPs. The anti-(SP) antibodies and antagonists may also be used to down regulate expression and activity of SP and as diagnostic agents for detecting the presence of SPs in sample. The disorders include for example: immune/autoimmune diseases (e.g. HIV, human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac

CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation.

Query Match 6.2%; Score 34.4; DB 22; Length 3943;
Best Local Similarity 53.0%; Pred. No. 4.6;
Matches 71; Conservative 1; Mismatches 62; Indels 0; Gaps 0;

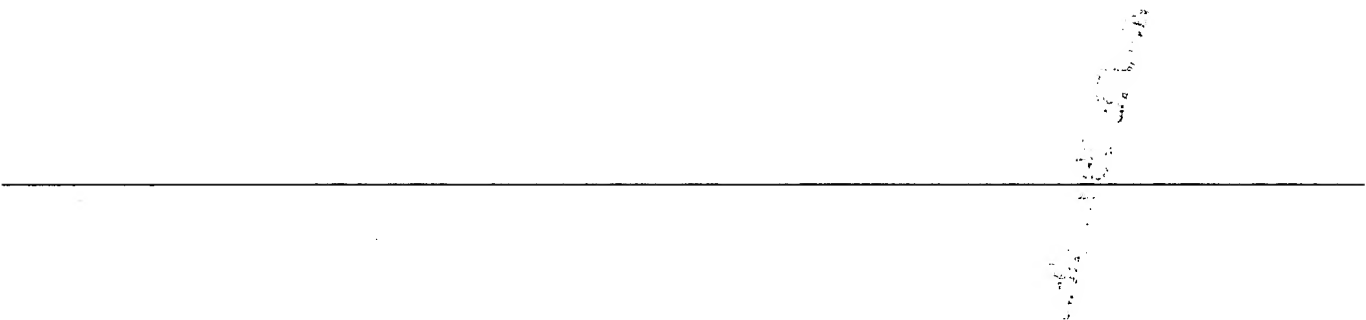
```
OY      98   tcaacttcagatatcctgtagacggatgltacgcagaagccccaataagatgagacaatatc 157
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB     1087  TYCATTTCATTTTTCACGGGAGAAAATFACATRGCAAGAGCTGAACNGTGAGGCCAAGACC 1028

OY     158   taattgaacaacgacctgatcgctgtgaaaaaacaaatttgatlacgysnaaccataaaggcatcaa 217
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    1027  TAATCAATAAAATGCGTAGACACTCAATAATATGTAAAGCTTTCTTAATGCAATAAGTACTTAA 968

OY     218   caggaacttgcgt 231
        ||| | | | | | | |
DB     967   CAGAGACTTGATAT 954
```

Search completed: June 27, 2002, 20:21:50
Job time: 6408 sec

[illegible]



THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 17:35:57 ; Search time 103.12 Seconds
(Without alignments)
1322.019 Million cell updates/sec

Title: US-09-727-769a-5
Perfect score: 555
Sequence: 1 ttgscgagtgtaattcctga.....atgtatccagctgtgattc 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfilas1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	344.2	62.0	555	4	US-09-324-910-5
2	344.2	62.0	1080	4	US-09-324-910-11
3	35.6	6.4	2489	3	US-09-141-047-7
4	34.8	6.3	1751	4	US-09-004-393B-3
5	34.2	6.2	3666	2	US-08-682-517-13
6	34.2	6.2	3666	2	US-08-682-517-14
7	34.2	6.2	4197	2	US-08-682-517-7
8	34.2	6.2	4197	2	US-08-682-517-8
9	33.6	6.1	4223	4	US-09-541-782-5
10	32.8	5.9	5502	3	US-08-836-134-1
11	32.4	5.8	7286	3	US-09-331-581-3
12	32.4	5.8	7938	3	US-09-331-581-14
13	31.8	5.7	1189	4	US-08-936-165A-195
14	31.6	5.7	1810	3	US-08-946-026-19
15	31.2	5.6	657	3	US-08-721-979A-74
16	31.2	5.6	657	3	US-08-836-501-74
17	31.2	5.6	726	3	US-08-836-501-78
18	31.2	5.6	1050	3	US-08-836-501-76
19	31.2	5.6	1071	3	US-08-836-501-77
20	31	5.6	1892	2	US-08-933-750C-66
21	31	5.6	1892	3	US-09-234-613-66
22	31	5.6	87350	3	US-08-781-891-79
23	30.4	5.5	7218	1	US-08-232-463-14
24	30.2	5.4	1007	1	US-07-721-761A-37
25	30.2	5.4	1007	1	US-07-978-687-37
26	30.2	5.4	1007	5	PCT-US91-05801-37
27	30.2	5.4	1573	1	US-07-721-761A-36

C	28	30.2	5.4	1573	1	US-07-978-687-36	Sequence 36, Appl
C	29	30.2	5.4	1573	5	PCT-US91-05801-36	Sequence 36, Appl
C	30	29.6	5.3	3126	3	US-08-898-978-1	Sequence 1, Appl
C	31	29.4	5.3	443	1	US-08-592-406-13	Sequence 13, Appl
C	32	29.2	5.3	5433	3	US-08-929-329-1	Sequence 1, Appl
C	33	29	5.2	1418	1	US-08-391-615-7	Sequence 1, Appl
C	34	29	5.2	1823	1	US-08-145-995A-1	Sequence 1, Appl
C	35	29	5.2	1823	1	US-08-451-747-1	Sequence 1, Appl
C	36	29	5.2	1823	3	US-09-134-852-1	Sequence 1, Appl
C	37	28.8	5.2	5116	1	US-08-038-682-1	Sequence 1, Appl
C	38	28.8	5.2	5116	1	US-08-302-832-1	Sequence 1, Appl
C	39	28.8	5.2	5116	2	US-08-530-198-1	Sequence 1, Appl
C	40	28.8	5.2	5116	2	US-08-469-880-1	Sequence 1, Appl
C	41	28.8	5.2	5116	2	US-08-728-470-1	Sequence 1, Appl
C	42	28.8	5.2	5116	2	US-08-617-697-1	Sequence 1, Appl
C	43	28.8	5.2	5116	4	US-08-719-641-1	Sequence 1, Appl
C	44	28.8	5.2	7218	1	US-08-232-463-14	Sequence 14, Appl
C	45	28.4	5.1	1455	3	US-09-009-494-1	Sequence 1, Appl

ALIGNMENTS

RESULT	1
US-09-324-910-5	Sequence 5, Application US/09324910
Patent No. 6251651	
GENERAL INFORMATION:	
APPLICANT: Yamaguchi, Shotaro	
APPLICANT: Matsura, Akira	
TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE	
TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF	
FILE REFERENCE: 0-54362	
CURRENT APPLICATION NUMBER: US/09/324, 910	
CURRENT FILING DATE: 1999-06-03	
EARLIER APPLICATION NUMBER: HEI-10-173940	
EARLIER FILING DATE: 1998-06-04	
NUMBER OF SEQ ID NOS: 12	
SOFTWARE: Patentln Ver. 2.0	
SEQ ID NO 5	
LENGTH: 555	
TYPE: DNA	
ORGANISM: Chryseobacterium gleum	
US-09-324-910-5	
Query Match	62.0%; Score 344.2; DB 4; Length 555;
Best Local Similarity	76.7%; Pred. No. 5.4e-96;
Matches 421; Conservative	0; Mismatches 128; Indels 0; Gaps 0;
QY	7 agtgaattcctgtagtgcctaaattcttattcaataaagaatcagctc 66
DB	7 agtgaattcctgtagtgcctaaattcttattcaataaagaatcagctc 66
QY	67 tgggaattcctgtagtgcctaaattcttattcaataaagaatcagctc 126
DB	67 tgggaattcctgtagtgcctaaattcttattcaataaagaatcagctc 126
QY	127 tatgaagaagccataaagatgagcaaatcttaataaagaatcagctc 186
DB	127 tatgaagaagccataaagatgagcaaatcttaataaagaatcagctc 186
QY	127 tatgaagaagccataaagatgagcaaatcttaataaagaatcagctc 186
DB	127 tatgaagaagccataaagatgagcaaatcttaataaagaatcagctc 186
QY	187 caattgtaaggaagccataaagatgagcaaatcttaataaagaatcagctc 246
DB	187 caattgtaaggaagccataaagatgagcaaatcttaataaagaatcagctc 246
QY	247 caattgtaaggaagccataaagatgagcaaatcttaataaagaatcagctc 306
DB	247 caattgtaaggaagccataaagatgagcaaatcttaataaagaatcagctc 306
QY	307 gatccctactatttcaagcggctcctgtaacagatacagcagatgagaaagcttcgct 366
DB	307 gatccctactatttcaagcggctcctgtaacagatacagcagatgagaaagcttcgct 366

```
OY 367 aacacctctgcgatactgcacacccgttctccttcttaagttaactgacggaagtgttat 426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 aacacaagctggagctgcgtctgtatctcttcttccgaacataacagcaggaattttac 426
OY 427 tacagaatctcgaataattcttactctgtatgacaacaatctgataaccacgtgtta 486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 tacagaagctgcgcaggtcttactgtatgataaacaactatgtataccaattgtgta 486
OY 487 ctgaataatttcaactgtcttcccttccgaatgcttccacccgcagatgtaccagc 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 ttaacacatctctcctcccttccgaatgttctccttcccccagcaccagtgtagcaagc 546
OY 547 tgtgattt 555
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 tgtgattt 555

RESULT 2
US-09-324-910-11
; Sequence 11, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; FILE REFERENCE: 0-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; EARLIER FILING DATE: 1998-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Chryseobacterium gleum
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (61)..(462)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (463)..(1017)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61)..(1017)
; US-09-324-910-11

Query Match 62.0%; Score 344.2; DB 4; Length 1080;
Best Local Similarity 76.7%; Pred. No. 7.1e-96;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

OY 7 agtgaattctctgattgtgcctacatataattcttatcatcaataaagaatcagctt 66
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 agtgaattctctgattgtgcctacatataattcttatcatcaataaagaatcagctt 528
OY 67 tgcgttacctcgaagcgcctccacacatgacatcaatgataccgttagagagatgt 126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 tgcgaactctcagagctctcctctgtatcaccttccagataccggttgaagagatgt 588
OY 127 tatgacaagaccataaagatgagacaatacttaatgacaacaagcgtatgactgtgaaa 186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 589 tatgacaagcgtccacaataatgagacaatacttaatgacaacaagcgtatgactgtgaaa 648
OY 187 caattgtatagaagaaacctaaagcacaacaggaacttgcgtgtggtgcgtgagcctac 246
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 649 cagctgcataatgataatctgagaagcttctacagaacacatgctgtgtatcatgggtatat 708
OY 247 cagctgcataatgataatctgagaagcttctcggagtagaagaaataattat 306
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 709 cagtagcaattctgttagagcttcaaaaatgcttcaggaaattctgtgaaaaaagaatcata 768
```

```
OY 307 gatccttcaactatttcaacgcgttctgttaacagataacagacatggagaacgcgttgc 366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 769 gatccttcaactatttccagcgcgtctgttaacagatcttgcattgagagcgttcagtaac 828
OY 367 aacacctctgcgatactgcacacccgttctccttcttaagttaactgacggaagtgttat 426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 829 aacacaagctggagctgcgtctgtatctcttccgaacataacagcaggaattttac 888
OY 427 tacagaatctcgaataattcttactctgtatgacaacaatctgataaccacgtgtta 486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 889 tacagaagctgcgcaggttcttactgtatgataaacaactatgtataccaattgtgta 948
OY 487 ctgaataatttcaactgtcttcccttccgaatgcttccacccgcagatgtaccagc 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 949 ttaacacatctctcctcccttccgaatgttctccttcccccagcaccagtgtagcaagc 1008
OY 547 tgtgattt 555
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1009 tgtgattt 1017

RESULT 3
US-09-141-047-7/C
; Sequence 7, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
; FILE REFERENCE: D6143
; CURRENT APPLICATION NUMBER: US/09/141,047A
; EARLIER FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 7
; LENGTH: 2489
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; LOCATION: -340..2149
; OTHER INFORMATION: Nucleotide sequence of gene encoding 120kDa
; OTHER INFORMATION: Immunoreactive protein
; US-09-141-047-7

Query Match 6.4%; Score 35.6; DB 3; Length 2489;
Best Local Similarity 45.9%; Pred. No. 0.21;
Matches 122; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

OY 276 tgcctcggagtaacgaaataatgatccttcaactatttccaagcgtccgt 335
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1843 TTCTTCCTTACTAGTTTCAGATCTTTTCTCCAACTTCACCTGATGATGATCTTACACT 1784
OY 336 aacagatacagcagatggagaaacgcttgcgttaaacacctctgcgagatctgcaccttcc 395
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1793 ACCACTCTACAGAGGTTTCGAATCTTCCGCTTACTTCAGAGATATTTCTCTTACT 1724
OY 396 cctctatcctaactgtagaagaatgttatgaagaagtcctagtaattcttacctgta 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1723 AGTTTCAGATCTTTTCTCCAACTTCAGATGATGATGATGATGATGATGATGATGATGAT 1664
OY 456 tgacaacaatctgataacataccaactgttactgactaactaatttcaactgtcttccgagtg 515
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1663 AGGTTGCAAACTGTTGCTTAACTTCAGAGATATTTCTCTTACTTACAGATGATGATGATGAT 1604
OY 516 ttctcctcaactgcacccgagatgat 541
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1603 TTTTCTCCAACTTCACCTGATGATGAT 1578

RESULT 4
US-09-004-393B-3
```

```

: Sequence 3 Application US/09004393B
: Patent No. 6310271
: GENERAL INFORMATION:
: APPLICANT: Hanson D., Andrew
: APPLICANT: Rathnassabapathi, Bala
: APPLICANT: Burnet, Michael
: TITLE OF INVENTION: Polynucleotides Encoding Choline Monoxygenase and
: FILE OF INVENTION: Plants Transformed Therewith
: FILE REFERENCE: UF-162
: CURRENT APPLICATION NUMBER: US/09/004,393B
: PRIOR APPLICATION NUMBER: 1998-01-08
: PRIOR FILING DATE: 1997-01-08
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 3
: LENGTH: 1751
: TYPE: DNA
: ORGANISM: Beta vulgaris
: US-09-004-393B-3

```

Query Match	6.3%	Score 34.8;	DB 4;	Length 1751;
Best Local Similarity	48.1%;	Pred. No. 0.32;		
Matches	99;	Conservative	0;	Mismatches 107;
			Indels	0;
			Gaps	0;

[illegible]

```

RESULT      5
US-08-682-517-13
: Sequence 13, Application US/08682517
: Patent No. 5874267
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Expression of surface layer proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,517
: FILING DATE:
: CLASSIFICATION:
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3666 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-682-517-13

```

Query Match	6.2%	Score 34.2;	DB 2;	Length 3666;
Best Local Similarity	47.4%;	Pred. No. 0.66;		
Matches 102;	Conservative	0;	Mismatches 113;	Indels 0;
			Gaps	0

Oy	272	aatlgcttcgcgagatlaacggaaaagaatatgatccctactactatttcaaggcgc	331
Dd	1013	AAGTACTGATGAAGTAGACGCTTTATGCACACTGGTGATGTCTTAATTCTC	1072
Oy	332	ctgtacaacagatacacgaatgagaaagcctgcgttaaacaccctctgcagatcgtac	391
Dd	1073	ATGTAATTCGTGGGGGTATGATNCAATTCCTTCAAGTTSAACAATACTACAGSTGCTT	1132
Oy	392	tttccctcttagctaactactgcagaaagtlltatlacaaagtcctcglaattctaac	451
Dd	1133	TTAATAATGGTGCAAAACAAAACCTTACAAAGTTACTATTAATAAACCCCTA	1192
Oy	452	tgtatgacaacacatcgtatafaccaactgbita	486
Dd	1193	CAGAAGCAAACAAACACTTTAAATGTTGGTTTGGTA	1227

```

      RESULT      6
US-08-682-517-14
; Sequence 14, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
; US-08-682-517-14

```

[illegible]

RESULT 7
US-08-682-517-7
; Sequence 7, Application US/08682517
; Patent No. 5874267

```
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682.517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-682-517-7
```

```
Query Match 6.2%; Score 34.2; DB 2; Length 4197;
Best Local Similarity 47.4%; Pred. No. 0.7;
Matches 102; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
```

```
Oy 272 aaagcttcgcgagtaacggaanaagaattatgacctcactatttcaagcggtc 331
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1197 AAGGACTGATGATGAAGTAAGTCTGTTATGCAACGTGATGCTTTCTAATTTCTACTGTTGTT 1256

Oy 332 ctgtacagatacagcatgagaaacgcttgcgttaacaccccttcgagatcgatccg 391
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 ATGTATCTGGGGTGTAGATACAAATTTCTTCAAGTGAAGAACTAAGTACAGTGTCTGAG 1316

Oy 392 ttctcttatagtctaatactgcaggaatggttatatcacagaagtcctgtaattcttacc 451
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1317 TTAATTAATGGTGCAAAACAAACTTACAAAGTTATATAAAACCCCTAAACTGGTAAAC 1376

Oy 452 tgaatgacacaatctgataccaataccaactgtgta 486
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1377 CAGAAGCAAAACAAACATTTAATGTTGTTTGTGA 1411
```

```
RESULT 8
US-08-682-517-8
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682.517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..184
; US-08-682-517-8
```

```
Query Match 6.2%; Score 34.2; DB 2; Length 4197;
Best Local Similarity 47.4%; Pred. No. 0.7;
Matches 102; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
```

```
Oy 272 aaagcttcgcgagtaacggaanaagaattatgacctcactatttcaagcggtc 331
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1197 AAGGACTGATGATGAAGTAAGTCTGTTATGCAACGTGATGCTTTCTAATTTCTACTGTTGTT 1256

Oy 332 ctgtacagatacagcatgagaaacgcttgcgttaacaccccttcgagatcgatccg 391
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 ATGTATCTGGGGTGTAGATACAAATTTCTTCAAGTGAAGAACTAAGTACAGTGTCTGAG 1316

Oy 392 ttctcttatagtctaatactgcaggaatggttatatcacagaagtcctgtaattcttacc 451
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1317 TTAATTAATGGTGCAAAACAAACTTACAAAGTTATATAAAACCCCTAAACTGGTAAAC 1376

Oy 452 tgaatgacacaatctgataccaataccaactgtgta 486
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1377 CAGAAGCAAAACAAACATTTAATGTTGTTTGTGA 1411
```

```
RESULT 9
US-09-541-782-5/c
; Sequence 5, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4223
; TYPE: DNA
; ORGANISM: S.pombe
; US-09-541-782-5
```

```
Query Match 6.1%; Score 33.6; DB 4; Length 4223;
Best Local Similarity 52.1%; Pred. No. 1.1;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
```

```
Oy 336 aacagatacagcatgagaaacgcttgcgttaacaccccttcgagatcgatccggttc 395
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3073 AGCATCAATATGCTTACGCACACCATATGATGCTCTTCTGTGATTTCTTCAAGTATG 3014

Oy 396 ctctaatactatctgcaggaatggttatatcacagaagtcctgtaattcttaccgta 455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3013 CTGTAATTAATTAATGATGATCTTAATTCCTGAACAAGTTGTGGTATGATTTCTCATGTC 2954

Oy 456 tgaacaacatctgataccaataccaactgtgta 479
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2953 TTTCTGCAATTCATCAATTCGGA 2930
```

```
RESULT 10
US-08-836-134-1/c
; Sequence 1, Application US/08836134A
; Patent No. 6020127
```

GENERAL INFORMATION:
APPLICANT: Mackenzie, Alex E.
APPLICANT: Korneiluk, Robert G.
APPLICANT: Mahadevan, Mani S.
APPLICANT: Mclean, Michael
APPLICANT: Roy, Natalie
APPLICANT: Ikeda, John-e
TITLE OF INVENTION: Neuronal Apoptosis Inhibitor Protein, Gene Sequence and
Patent No. 6020127
FILE REFERENCE: Mutations Causative of Spinal Muscular Atrophy
CURRENT APPLICATION NUMBER: US/08/836,134A
CURRENT FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 5502
TYPE: DNA
ORGANISM: Homo sapiens
US-08-836-134-1

Query Match 5.9%; Score 32.8; DB 3; Length 5502;
Best Local Similarity 58.0%; Pred. No. 2.1;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 238 tggagctaccacgttgtaaatatgtgtaagctataaaatgcttcggaagtaacggaataa 297
Db 3805 TGGATCATATATTCTTCATATGTTGAGACTTGGAAATCTTCAGGAATGACTGAATAA 3746
Qy 298 agaatatgacccctcaccatatttcaagcgcgtctgttaa 337
Db 3745 ACATTATATTGCCCTCCAGATCCACAGACAGTCTTTCA 3706

RESULT 11
US-09-331-581-3
Sequence 3, Application US/09331581
Patent No. 6130070
GENERAL INFORMATION:
APPLICANT: TOMODA, Hideki
APPLICANT: HAMA, Yuko
APPLICANT: KUWAGAI, Hiromichi
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
FILE REFERENCE: 0059-1142-0PCT
CURRENT APPLICATION NUMBER: US/09/331,581
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/JP98/04929
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: JP 9-314608
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3
LENGTH: 7286
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3

Query Match 5.8%; Score 32.4; DB 3; Length 7286;
Best Local Similarity 48.4%; Pred. No. 3.1;
Matches 90; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 117 agacgagtgatgcaagagccataagatgagacaatcttatgacaagcgctatga 176
Db 4574 agatgctctatgtagctgtaaaataaataagacgtccacaagaaaaaataaaaaa 4633
Qy 177 ctgtgaaaaacaatttatacgaaacctaaggaagcaacaggaacttgctgtgyc 236

Db 4634 aaagtaccttcgagcg 4693
Qy 237 gtgagctaccacggttgcaatattgtaagctataaaatgcttcgaggaagcaaaa 296
Db 4694 aaagaaccagcgatccacatgataagatacatatgagtttggacaacaccacaac 4753
Qy 297 aagaat 302
Db 4754 tagaat 4759

RESULT 12
US-09-331-581-14
Sequence 14, Application US/09331581
Patent No. 6130070
GENERAL INFORMATION:
APPLICANT: TOMODA, Hideki
APPLICANT: HAMA, Yuko
APPLICANT: KUWAGAI, Hiromichi
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
FILE REFERENCE: 0059-1142-0PCT
CURRENT APPLICATION NUMBER: US/09/331,581
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/JP98/04929
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: JP 9-314608
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 7938
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14

Query Match 5.8%; Score 32.4; DB 3; Length 7938;
Best Local Similarity 48.4%; Pred. No. 3.2;
Matches 90; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 117 agacgagtgatgcaagagccataagatgagacaatcttatgacaagcgctatga 176
Db 5226 agatgctctatgtagctgtaaaataaataagacgtccacaagaaaaaataaaaaa 5285
Qy 177 ctgtgaaaaacaatttatacgaaacctaaggaagcaacaggaacttgctgtgyc 236
Db 5286 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaagaatccttcgagcg 5345
Qy 237 gtgagctaccacggttgcaatattgtaagctataaaatgcttcgaggaagcaaaa 296
Db 5346 aaagaaccagcgatccacatgataagatacatatgagtttggacaacaccacaac 5405
Qy 297 aagaat 302
Db 5406 tagaat 5411

RESULT 13
US-08-936-165A-195
Sequence 195, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard

```

1  APPLICANT: Pratt, Julie
2  APPLICANT: Reichard, Richard
3  APPLICANT: Rosenberg, Martin
4  APPLICANT: Ward, Judith
5  TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
6  TITLE OF INVENTION: Polypeptides and Their Uses
7  NUMBER OF SEQUENCES: 534
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: SmithKline Beecham Corporation
10 STREET: 709 Swedeland Road
11 CITY: King of Prussia
12 STATE: PA
13 COUNTRY: USA
14 ZIP: 19406-0939
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Diskette
17 COMPUTER: IBM Compatible
18 OPERATING SYSTEM: DOS
19 SOFTWARE: FASTSEQ for Windows Version 2.0
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/936,165A
22 FILING DATE: 24-SEP-1997
23 CLASSIFICATION: 536
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: 60/027,032
26 FILING DATE: 24-SEP-1996
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Gimm1, Edward R
29 REGISTRATION NUMBER: 38,891
30 REFERENCE/DOCKET NUMBER: P50549
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: 610-270-4478
33 TELEFAX: 610-270-5090
34 TELEX:
35 INFORMATION FOR SEQ ID NO: 195:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 1189 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: single
40 TOPOLOGY: linear
41 MOLECULE TYPE: Genomic DNA
42 ;
43 US-08-936-165A-195

```

Query Match 5.7%; Score 31.8; DB 4; Length 1189;
Best Local Similarity 56.1%; Pred. No. 2.3;
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0.

OY 159 aatgacaaagcgctatgcattgaaaaacaattglatacggaaaactaaagcatcaac 218
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 726 AATGACAGCGCCCTTAACTGTGATGAACAAGTTATTGCACAAAGCTTAACAGATAAAC 785

OY 219 aggaactgcgtctgtgcgttgagaactaccagttccaatatgtgtaa 265
+ ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 786 CAGTTAAAGTATGTTAACTGACCTGTAACAATTCTAAACGTGTCA 832

RESULT 14
US-08-946-026-19
Sequence 19, Application US/08946026
Patent No. 6034218
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Dillon, David L.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Mitcham, Jennifer L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: seed and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: Usa
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,026
FILING DATE: 07-OCT-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Mak1, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.424C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ. ID NO.: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1810 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-946-026-19

	Query Match	Best Local Similarity	Score	DB	Length
	5.7%	54.2%	31.6	3	1810
			Pred. NO.	3.1	
	Matches	Conservative	0	Mismatches	54
				Indels	0
				Gaps	0
Qy	97	atccattcagatgataccggtatgacgagtgattatgaaagagcccatagaatgtgacacaatc	156		
Db	1419	AACGAGCAGCAAGAAATCATCTCAGTTTATTAATAATCAAGAGACATCTTCAAGCAAAAAGC	1478		
Qy	157	ttaatgaacaaacgcgtactgactgtgaaaacaatttgtatatacggaaacctaagcat	214		
Db	1479	ATCATTTATATGAGGCCAGATGACTCTGAAGAAACAATTTAAAAAAGCCATCCAGAGAAAT	1536		

RESULT 15
 US-08-721-979A-74
 : Sequence 74, Application US/08721979A
 : Patent No. 6113911
 :
 GENERAL INFORMATION:
 :
 APPLICANT: Bliz, Hans
 APPLICANT: N'Guyen, Ngoc Thien
 APPLICANT: Baussant, Thierry
 APPLICANT: Trudel, Michel
 TITLE OF INVENTION: PEPTIDE FRAGMENT OF RESPIRATORY
 TITLE OF INVENTION: SYNCTIAL VIRUS PROTEIN G, IMMUNOGENIC AGENT, PHARMACEUTICAL
 TITLE OF INVENTION: COMPOSITION CONTAINING IT AND PREPARATION PROCESS
 NUMBER OF SEQUENCES: 75
 :
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Gordon W. Hueschen
 STREET: 715 The "H" Bldg., 310 East Michigan
 STREET: Avenue
 CITY: Kalamazoo
 STATE: MI
 COUNTRY: USA
 ZIP: 49007
 :
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 :
 SOFTWARE:
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/721,979A
 FILING DATE: October 4, 1996
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 16:25:12 ; Search time 2534.16 Seconds
(Without alignments)
2955.934 Million cell updates/sec

Title: US-09-727-769a-5

Perfect score: 555
Sequence: 1 ttggcgcagtgtaattccctga.....atgtatccagctgtgattt 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estda: *
2: em_esthm: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hlc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hlc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vtc: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	40.8	7.4	606	12	FR0009895
2	40.4	7.3	233	9	AA767965
3	39	7.0	532	10	BI893381
4	39	7.0	570	9	AW937387
5	39	7.0	570	9	AW937408
6	37.6	6.8	858	10	BF697486
7	37.4	6.7	531	10	BM092866
8	37.2	6.7	728	12	AG164728
9	36.8	6.6	550	12	AQ982897
10	36.6	6.6	555	12	BG683527
11	36.4	6.6	491	12	AQ143297
12	36.2	6.5	668	10	BI183607
13	35.6	6.4	642	12	AZ525610
14	35.6	6.4	686	12	AQ390063
15	35.6	6.4	1101	12	CNS0100X
16	35.4	6.4	625	9	AT722559
17	35.4	6.4	894	12	AZ531159

C	18	35.4	6.4	938	12	CNS0067J	AL065906 Drosophila
C	19	35.2	6.3	635	12	AQ982969	AQ982969 RPI-23-2
C	20	35	6.3	488	10	BI366489	BI366489 RE52036.5
C	21	35	6.3	539	12	AQ708947	AQ708947 HS_2121_A
C	22	35	6.3	666	10	BI240734	BI240734 RE37648.5
C	23	35	6.3	687	9	AV682345	AV682345 AV682345
C	24	35	6.3	1001	12	CNS012LO	AL101718 Drosophila
C	25	34.8	6.3	367	10	BE581125	BE581125 K942F01.Y
C	26	34.8	6.3	434	10	BE579657	BE579657 K931C03.Y
C	27	34.8	6.3	455	10	BE579625	BE579625 K931E10.Y
C	28	34.8	6.3	590	12	AZ285039	AZ285039 RPI-23-1
C	29	34.8	6.3	600	12	AZ005132	AZ005132 RPI-23-3
C	30	34.8	6.3	685	12	AQ995355	AQ995355 RPI-23-2
C	31	34.8	6.3	759	9	AI541345	AI541345 pccl.2-7.
C	32	34.6	6.2	533	12	AQ413419	AQ413419 RPI-11-1
C	33	34.4	6.2	271	12	BI188461	BI188461 036_E_02-
C	34	34.4	6.2	271	12	CNS07SH2	AL625400 T3 end of
C	35	34.4	6.2	567	9	AW398915	AW398915 EST309415
C	36	34.2	6.2	501	12	AZ905458	AZ905458 RPI-24-1
C	37	34.2	6.2	978	12	CNS01POH	AL155145 Anopheles
C	38	34	6.1	156	9	AW164768	AW164768 se77605.Y
C	39	34	6.1	351	9	AW216307	AW216307 687047607
C	40	34	6.1	448	12	AZ802912	AZ802912 2M063D04
C	41	34	6.1	477	12	BH634370	BH634370 SALK_0452
C	42	34	6.1	564	10	BF296339	BF296339 035PBF05
C	43	34	6.1	571	10	BM023631	BM023631 1e82H05.Y
C	44	34	6.1	664	12	AQ356929	AQ356929 CTBR-E1-
C	45	34	6.1	846	10	BM398346	BM398346 5009-0-44

ALIGNMENTS

RESULT 1
LOCUS FR0009895/C
DEFINITION F.rubripes GSS sequence, clone 032C09ad10, genomic survey sequence.
ACCESSION AL001177
VERSION AL001177.1 GI:2439029
KEYWORDS GSS: genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes.

REFERENCE
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranta,Y.,
Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk

COMMENT
Vector: Bluescript II KS
V-type: phagemid
PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
source location/Qualifiers
1. 606
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_11b="cosmid 032C09"
/clone="032C09ad10"

BASE COUNT 165 a 150 c 110 g 160 t 21 others
ORIGIN

Query Match 7.4% Score 40.8; DB 12; Length 606;
Best Local Similarity 50.3% Pred. No. 0.63;
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
Gy 183 aaacaatttgatcaagaaacctaagcgtcaacggaacttgctgtgagcgag 242

```

Db      553 AGAATAATGTATGTGCTNACAGAGATGATCCACATTTTNGAACCTNTTGGT 494
Oy      243 ctcaacagcttgaatcgtatgtaagctataaaatgctccggaagtaacggaataaagat 302
Db      433 AAGGATGCTGCTGAAAGTCGAAAGCTAAATTAATTAATCTATATGCGCAAGTNGAAT 434
Oy      303 tatgatccctcactatttccaagcgctctgtaacagatacagaatggaagacgcttg 362
Db      433 TTATACCTCATCTTATNTGAAATCTCTGTAGGAAGTAGATGCTGATAGCGCGCG 374
Oy      363 cgcta 367
Db      373 CTGTGA 369

RESULT 2
AA767965 233 bp mRNA linear EST 07-FEB-1998
LOCUS    nv96a12.s1 NCI.CGAP_pri16 Homo sapiens cDNA clone IMAGE:1237630,
DEFINITION
RNA sequence.
ACCESSION AA767965
VERSION    AA767965.1 GI:2818980
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 233)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgaabs-remail.nih.gov
          Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,
          M.D., Michael R. Emmert-Buck, M.D., Ph.D.
          cDNA Library Preparation: David B. Kitzman, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          www-bio.llnl.gov/dbp/image/image.html
          Insert Length: 332 Std Error: 0.00
          Seq primer: -40m13 fwd. ET from Amersham.
FEATURES
Source    Location/Qualifiers
          1..233
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_image="1237630"
             /clone_lib="NCI.CGAP_pri16"
             /sex="male"
             /tissue_type="tumor"
             /lab_host="DH10B"
             /note="Organ: prostate; Vector: pAMP10; mRNA made from
             invasive prostate tumor cells; cDNA made by oligo-dT
             priming. Non-directionally cloned. Size-selected on
             agarose gel, average insert size 600 bp. Library made by
             D. Kitzman, NIH."
BASE COUNT      84 a      41 c      40 g      68 t
ORIGIN
Query Match      7.3%; Score 40.4; DB 9; Length 233;
Best Local Similarity 49.5%; Pred. 0.59;
Matches 104; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
Oy      18 tgatgtagcatcaataatcttattcaatcaataagaatcagctctgcgtactc 77
Db      21 TGGTTGTTTAACATACATACAGTTCACATATGACATTTACTGTTAGGCGTC 80
Oy      78 taagcgccctcacatcacatcagatatactgtaagcgatgttaagcaagagc 137

```

```

Db      81 TTTTGTCCAAAACAGTCGGCTTCCATACAGCAGCCTCTGCACATGTAGCTGAGC 140
Oy      138 cataagatgaacaacacttaatagaacaagcgtactgtaaaaaaatgtata 197
Db      141 AGATAAATATGCGCTAGTCCAAATTAAGTAGCTGTAGTATAAATGCAAGTGCAT 200
Oy      198 cggaaacctaaagcatcaacagacttg 227
Db      201 TCCAAAACAGTATTCAAAAAAGATCG 230

RESULT 3
BI893381 532 bp mRNA linear EST 30-NOV-2001
LOCUS    sa165d11.y1 Gm-cl068 Glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION
ID: Gm-cl068-3621 5', mRNA sequence.
ACCESSION BI893381
VERSION    BI893381.1 GI:16105641
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 532)
AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corvelli,V., Khanna
          A., Bolla,B., Marra,M., Hillier,L., Kucab,T., Martin,J., Beck,C.,
          Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers
          T., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
          R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann
          R., Waterston,R. and Wilson,R.
          Public Soybean EST Project
          Unpublished (1999)
TITLE     Public Soybean EST Project
JOURNAL   Unpublished (1999)
COMMENT   Contact: Shoemaker R/Public Soybean EST project
          Public Soybean EST Project
          Washington University School of Medicine
          444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@wustl.wustl.edu
          This clone is available through: Resgen, Invitrogen Corp. 2130
          South Memorial Parkway Hunttsville, AL 35801 For further information
          call: (800)-533-4363 or contact via email: ccu@resgen.com
          Seg primer: -40RP from Gibco
          High quality sequence stop: 425.
FEATURES
Source    Location/Qualifiers
          1..532
             /organism="Glycine max"
             /db_xref="taxon:3847"
             /clone_image="GENOME SYSTEMS CLONE ID: Gm-cl068-3621"
             /clone_lib="Gm-cl068"
             /tissue_type="leaf, drought stressed, 1 month old plants,
             greenhouse grown"
             /lab_host="DH10B"
             /note="Vector: pluescript II SK+, Site-1: EcoRI, Site-2:
             XhoI; The cDNA library was constructed from mRNA isolated
             from drought stressed leaf tissue of the cultivar Williams
             82. The month old greenhouse grown plants were deprived of
             water for 3 days prior to harvesting the stressed leaf
             tissue. Complementary DNA was synthesized from mRNA using
             a primer consisting of a poly(dT) sequence with a XhoI
             restriction site. EcoRI adapters were ligated to the
             blunt-ended cDNA fragments followed by XhoI digestion. The
             cDNA fragments were directionally cloned into the
             EcoRI-XhoI restriction site of the pluescript vector. The
             ligated cDNA fragments were transformed into DH10B host
             cells (GibcoBRL). This library was constructed in the
             laboratory of Dr. Randy Shoemaker."
BASE COUNT      160 a      115 c      112 g      145 t
ORIGIN

```

Query Match	7.0%	Score 39	DB 10	Length 532
Best Local Similarity	68.4%	Pred. No. 2		
Matches 54	Conservative 0	Mismatches 25	Indels 0	Gaps 0
Oy	34	aattcttattcaatcaataaagaatcagcttcgtgtaacctcagcgcgtccaca	93	
Db	5	AATCTTTTGTTCACAAACAAACGATGATTTGGTGGCTTCTTCTACGAGATGTCACCA	64	
Oy	94	tgcataccatccagatc	112	
Db	65	TGAATCATCTTCAGTTACC	83	
RESULT 4				
AM937387/c				
LOCUS	AM937387	570 bp	mRNA	linear
DEFINITION	QV3-DT0043-180100-061-c07 DT0043	Homo sapiens	CDNA,	mRNA sequence.
ACCESSION	AM937387			
VERSION	AM937387.1	GI:8112810		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Enkavirota, M.; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Cararrhina; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 570)			
	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Grudman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunslein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.			
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7)	3491-3496	(2000)
MEDLINE	20202663			
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4t2-QV3-DT0043-180100-061-c07&f3=2000-01-18&t4=1) Seq primer: puc 18 forward High quality sequence stop: 569.			
FEATURES				
source	Location/Qualifiers			
	1..570			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_lib="DT0043"			
	/dev_stage="Adult"			
	/note="Organ: denis_drahs; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
	low stringency conditions."			
BASE COUNT	203 a	83 c	94 g	190 t
ORIGIN				
Query Match	7.0%	Score 39	DB 9	Length 570
Best Local Similarity	52.8%	Pred. No. 2		
Matches 84	Conservative 0	Mismatches 75	Indels 0	Gaps 0
Oy	182	aaaacacatttctatcagcaaacctcaagagcgtacacaggaactcgcgtgagcgatg	241	
Db	480	AAAGAGATGTTTATAGGAGAAATCTGGTTACTTCTTGTGATATTAGACATTTATGGATTAAA	421	

QY	242	gctccacgcttgcatatggtlaagcctataaagtcctccggagtaacggaataaagaa	301
Db	420	gctacctgtttatattatggcggttagccttagaagctaccCAAAAGAAAGGTAATTTG	361
QY	302	ttatcgatcctcactatcttcaacgagtcctgtaacag	340
Db	360	CTAAAGAAATTTCAAAATTTAAAGTTGCTCTTTAAAGA	322
RESULT	5		
AM937408/c		570 bp	linear
LOCUS			EST 30-MAY-2000
DEFINITION	OV3-DT0043-190100-061-c07 DT0043 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	OV37408		
VERSION	AM937408.1	GI:8112831	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE			
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunsfeld, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed		
JOURNAL	sequence tags		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
COMMENT	20202663		
	Contact: Simpson A.J.G.		
	Laboratory of Cancer Genetics		
	Ludwig Institute for Cancer Research		
	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil		
	Tel: +55-11-2704922		
	Fax: +55-11-2707001		
	Email: asimpson@ludwig.org.br		
	This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL		
	(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-DT0043-190		
	100-061-007&tl3=2000-01-19&tl4=1)		
	Seq primer: puc 18 forward		
	High quality sequence stop: 569.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..570		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="DT0043"		
	/note="stage="Adult"		
	/dev="Organ: dens_drsch; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	203 a 83 c 94 g 190 t		
ORIGIN			
	Query Match	7.0%; Score 39; DB 9; Length 570;	
	Best Local Similarity	52.8%; Pred. No. 2;	
	Matches	84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;	
QY	182	aaanaacatctgcatagcaaacctaaagcatcaacaggaactgtctgtgctgtgca	241
Db	480	AAAAGAGTGTATATAGGAAATATCGTCTTATATTAAGAACTTATAGGATTTAA	421
QY	242	gtcacacgcttgcaatatctgtttaagctataaagtcctccggagtaacggaataaagaa	301

Db 420 GCTACCTTGTTATTATATGGGGCTAGGCGTAAGTAGTACCCCAAGAAAAAGGTAATTTG 361

Qy 302 ttatgtacccctcaactatttcaagcggctcctgtacacg 340
|| || ||||| ||||| ||| |||

Db 360 CTAAGAATAATTTCACAAATAATTAAAGTGTCTTTAAAG 322

RESULT	6
BP697486/c	
LOCUS	BP697486
DEFINITION	858 bp mRNA linear EST 22-DEC-2000
ACCESSTION	60212337F1 NIH_MGC_56 Homo sapiens CDNA clone IMAGE:4286394 5', mRNA sequence. bp607405

VERSION	BF697486.1	GI:11982894
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eultheria; Primates; Carnivora; Hominiidae; Homo. 1 (bases 1 to 858)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: ATCC
cDNA Library Preparation: CLOVETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM1122 row: m column: 19
High quality sequence stop: 537.

FEATURES	Location/Qualifiers
source	1. .858

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4286394"
 /clone_1ib="NH.MGC.56"
 /tissue_type="Primitive neuroectoderm"
 /lab_host="DH10 (T1 phage-resistant)"
 /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site:1
 SfiI (ggcgccttggcgc); Site:2: SfiI (ggcgccttggcgc).
 Double-stranded cDNA was prepared from cell line RNA. 5
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCATTAATGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTGAGAGCCGACGAGCGCCGACATC-dT(30)BN-3'
 where B = A, C, G or G and N = A, C, G, or T). Average
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 laboratories (Palo Alto, CA)"

BASE COUNT	247	a	168	c	151	g	292	t
ORIGIN								

Query Match	6.88;	Score 37.6;	DB 10;	Length 858;
Best Local Similarity	47.58;	Pred. No. 5.9;		

Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```

Oy 191 ttgtatacggaaacctaaagcattcaacaggaacttgcctgtgtgcygtgagctaccag 250
    ||||| - - - - - ||||| - - - - -
Db 314 ttgcttggaatttaacagtaataatctgaamaattagctataactccaggaaccata 255

```

251 ttgcacatattgtgaagctataaaatgccttcgagagtaacgaaaaagaattattgac 310

Db 254 ATGGTACAGTGGTAAATGCAAAAGTCCACCTGAATATAATTTCATCACTCCC 195

311 ctcaactatlttccaagcgcgtctgttaacagatacagcatgtgagaacgcttgcttaaca 370

Db 194 TTCAGAAATTCACCACTGTCTCATTTCTATTTCAGCTGAAGAAATAAGGACATAAA 135

```
OY      371 cctcctgcgagatcgcaaccggttccctctaagtacaaactgcagaataatttat 4208
```

|||||
||| |

```
Db     134 CCCATACCTTAATTGGAGCCTTTCGCTGAACGTATAATTAAATGAATGATTTT   79
```

RESULT	7
BM092866	
LOCUS	531 bp mRNA linear EST 30-NOV-2001
DEFINITION	sa020203.y1 Gm-cl065 glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl065-9269 5', mRNA sequence.	
ACCESSION	BM092866

VERSION	BM092866.1	GI:17021832
KEYWORDS	EST.	
SOURCE	soybean.	
ORGANISM	Glycine max	

Euanyulca, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae Glycine.

REFERENCE
1 (bases 1 to 531)

AUTHORS
Shoemaker, R., Kelm, P., Vochkin, L., Erpelidng, J., Corvett, V., Khanna

TITLE
Public Soybean EST Project

Unpublished (1999).
Contact: Shoemaker R/Public Soybean EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
high quality sequence stop: 422.

FEATURES
source

```

/organism="Glycine max"
/db_xref="taxon:3844"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1065-9269"
/clone_1id="Gm-c1065"
/tissue_type="germinating shoots"
/lab_host="DH10 $\alpha$ "
/notes="Vector: plasmid pUC19. Site 1: EcoRI; Site 2: XhoI. The cDNA library was constructed from mRNA isolated from germinating shoots of the cultivar Williams. The seeds were allowed to germinate for 24 hours prior to being cold stressed for 2 days at 4C. Complementary DNA was

```

poly(dT) sequence with a λ adapters were ligated to the

followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy

BASE COUNT	153 a	127 c	87 g	164 t
ORIGIN				

Query Match 6.78; Score 37.4; DB 10; Length 531;

Best Local Similarity	67.1%;	Pred. NO. 5.7/;							
Matches	53;	Conservative	0;	Mismatches	26;	Indels	0;	Gaps	0;

34 aatcttattcaatcaataagaatcagtccttgcggtacctctacgagctctaacca 93

Db 91 AATCTTTTGTCCAAACAACGATGAATTGGTCGTTTTTCTACGGAGTCGCACCA 15

```

QY 94 tgcatacattcagatc 112
Db 151 TGAATCATCTTCAGTTACC 169

RESULT 8
AG164728/c
LOCUS AG164728
DEFINITION Pan troglodytes DNA, clone: RP43-031P16.TJ, genomic survey
sequence.
ACCESSION AG164728
VERSION AG164728.1 GI:16694406
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
Male BAC Library clone:RP43-031P16.TJ.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totochi,Y., Watanabe,H. and Sakaki,Y.
JOURNAL BAC end sequences of Library RPCI-43
REFERENCE
AUTHORS Unpublished
2 (bases 1 to 728)
TITLE Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totochi,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpesgsc.riken.go.jp, URL:http://hsp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the Rad process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1. 728
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-031P16.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT 164 a 188 c 169 g 205 t 2 others
ORIGIN
Query Match 6.7%; Score 37.2; DB 12; Length 728;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 123 atgttatgcgaagccatgaatgagacaatcttaataaacaagcgcatgactgtga 182
Db 448 ATGTGATTCAGGAGCATGCTTAAGGACACACGCTTAATAGTGAGCTGTGGCAGCATAT 389
QY 183 aaacacatttgatacgaagaacaaagacatcaacaggaactgctgtgtgctgtgag 242
Db 388 TCAAGCTCAGGCTGTGGTGTAATATATATATGTCATTAATGTGTATGAGACTGTGTGAC 329
QY 243 ctaccacgttgcaaatatgtgtaagctataaatagtcttcggaqtaacggaanaaagaat 302
Db 328 CTAATACGGGTGCTCTGCTGTGAGCCAAAGTAATGGGTCCGATCTCACAGCCAAAGCTGT 269
QY 303 tatgtga 308
Db 268 TGGTGA 263

```

```

RESULT 9
AO982897
LOCUS AO982897
DEFINITION Mus musculus 550 bp DNA linear GSS 30-JAN-2000
, DNA sequence.
ACCESSION AO982897
VERSION AO982897.1 GI:6816102
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aklnret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
JOURNAL Mouse BAC End Sequences from Library RPCI-23
COMMENT Unpublished (1999)
Other GSS: RPCI-23-262D23.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Ressea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac.ends/mouse/bac_end_intro.html
Plate: 262 row: D column: 23
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 550
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-262D23"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1;
EcoRI; Site:2; EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 181 a 122 c 109 g 137 t 1 others
ORIGIN
Query Match 6.6%; Score 36.8; DB 12; Length 550;
Best Local Similarity 53.9%; Pred. No. 8.6;
Matches 97; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 40 ttatcaatcaataaagaatcagcttcgtgtaaccctcaagcgctccatccatgcatc 99
Db 201 TTTAAACAACAACAACAACAACAACAAGTATGATTCACAGCAGCAGCATCGAAGC 260
QY 100 acattcagatatacctgtagacggaatgtatgcaagagccatgaatgagacaatctta 159
Db 261 ACACAAACAACAATGGTGTGGCAGAGTGTGTAATAAAGAGAGATTAAGACAACAATAAGC 320
QY 160 atgaacaacgcatatga-cgtgtgaaacaacattgtatacgaagaacctaagcatcaac 218
Db 321 CAGTAATGTGCTATCTTCTTTAAAGAACCTTAATATGAGAAATGGCATGACATCATC 380

```

```

RESULT 10
LOCUS      BG683527/c      555 bp      mRNA      linear      EST 01-MAY-2001
DEFINITION 60265138771 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4762196 3',
            mRNA sequence.
ACCESSION  BG683527
VERSION     BG683527.1
KEYWORDS    EST.
            GI:13914924
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 555)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
JOURNAL     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
COMMENT     Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: L10M1615 row: n column: 21
            High quality sequence start: 37
            High quality sequence stop: 555.
            Location/Qualifiers
                1..555
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_image="IMAGE:4762196"
                /clone_id="NIH_MGC_47"
                /tissue_type="neuroblastoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCACGAC(g). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."

BASE COUNT      160 a      116 c      97 g      182 t

Query Match      6.6%; Score 36.6; DB 10; Length 555;
Best Local Similarity 50.3%; Pred. No. 9.8;
Matches 90; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Y 339 agatcacgacatgaggaacgtctggttaaacaccccttcgagatcgatcgttctc 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 AATATACACGACGAAACGCTATTATGATCCCTTCAGTGATATTCCTCTGTTCTCC 227

Y 399 ttatgctaatactcgaggaatgtttatacagaagtcctagtaattcctaactgata 458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 TAATGCTCAAAATCATMAGGAGCCTCTGAAATGATGATCTGCTAAATGATAGACATCTGG 167

Y 459 caacaatctgatacaatacaactgtgtactgaactaaattcctaactgcttccgagatgt 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 CATCACTGTCAGACACTGTAACACCTTCAGTCTCAGTCTGTGCAATTAATCTCTGTTCTT 108

RESULT 11
LOCUS      AQ143297      491 bp      DNA      linear      GSS 24-SEP-1998
DEFINITION HS-3067_B1_A11.MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3067 Col=21 Row=B, DNA sequence.
ACCESSION  AQ143297
VERSION     AQ143297.1
KEYWORDS    GI:3533950

```

```

KEYWORDS     GSS.
SOURCE        human.
ORGANISM      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 491)
AUTHORS       Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
            Hood,L.
TITLE         Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL       Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE       99380589
COMMENT       Contact: Mahairas GC, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3067 row: B column: 21
            Class: BAC ends
            High quality sequence stop: 491.
            Location/Qualifiers
                1..491
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_image="Plate=3067 Col=21 Row=B"
                /clone_id="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="Organ: sperm; Vector: pBelobAC11; BAC clones in
                E-Coli DH10B"

BASE COUNT      169 a      93 c      90 g      131 t      8 others

Query Match      6.6%; Score 36.4; DB 12; Length 491;
Best Local Similarity 50.3%; Pred. No. 11;
Matches 88; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Y 152 aaattcattgaacaacgctatgactgtgtaaaaaaatctgtatagcgaacctaaagg 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 AACACCAATGTGATTCCTGGAAGGACATGCAGATTAACATACATGACTACGCTG 375

Y 212 catcacaggaactgctgtgtggtgagctaccacgctgcacaatctggttaagcata 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 AACACCAATGTGATTCCTGGAAGGACATGCAGATTAACATACATGACTACGCTG 435

Y 272 aaatgcttcgaggaatacaaaaagaatattatgactcttcaactttccaag 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 ACAATGCTCTTAACAGACCCATGACCAAGATGATGATCTGATTAACACGCTCAAG 490

RESULT 12
LOCUS      BI183607      668 bp      mRNA      linear      EST 10-JUL-2001
DEFINITION UNI-P-FN-ca-e-10-0-UNI.s1 UNI-P-FN Sus scrofa cDNA clone
            UNI-P-FN-ca-e-10-0-UNI 3', mRNA sequence.
ACCESSION  BI183607
VERSION     BI183607.1
KEYWORDS    GI:14658016
SOURCE      pig.
ORGANISM     Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE     1 (bases 1 to 668)
AUTHORS       Caetano,A.R., Johnson,R.K., and Pomp,D.
            Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
            Unpublished (2001)
JOURNAL       Contact: Pomp, D
            Department of Animal Science

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

```
Run on:      June 27, 2002, 20:15:32 ; Search time 84.27 Seconds
              (without alignments)
              243.843 Million cell updates/sec
```

Title: US-09-727-769A-6
Perfect score: 991
Sequence: I LASVIPDYATLNSLNFQIKN.....FSLISGCSPPSPADVSSCGF 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Maximum Match 100%
Listing first 45 summaries

Database :
A_Geneseq_032802 : *

1:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT *
2:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT *
3:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT *
4:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT *
5:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT *
6:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT *
7:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT *
8:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT *
9:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT *
10:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT *
11:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT *
12:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT *
13:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT *
14:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT *
15:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT *
16:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT *
17:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT *
18:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT *
19:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT *
20:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT *
21:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT *
22:	/SIDS/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	991	100.0	185	22	AA084386	Amino acid sequenc
2	991	100.0	320	22	AA084387	Amino acid sequenc
3	874	88.2	185	21	AA044582	Chryseobacterium g
4	874	88.2	319	21	AA044582	Chryseobacterium g
5	108	10.9	20	22	AA084385	Internal peptide c
6	96	9.7	20	22	AA084384	N-terminal peptide
7	91	9.2	594	21	AA052451	Arabidopsis thalia
8	91	9.2	594	21	AA052479	Arabidopsis thalia
9	85.5	8.6	477	22	AA030815	Arabidopsis thalia
10	85	8.6	264	21	AA025396	Amino acid sequenc
11	85	8.6	292	21	AA025395	Arabidopsis thalia

12	85	8.6	294	21	AABG25394	Arbidopsis thaliana
13	84	8.5	3726	22	ABBG63947	Drosophila melanog
14	82	8.3	927	20	AAV35054	Chlamydia pneumonia
15	82	8.3	928	20	AAAB88423	Chlamydia pneumonia
16	82	8.3	928	21	AAAY90239	Chlamydia antigen
17	81	8.2	270	21	AAAG21040	Arabidopsis thaliana
18	81	8.2	304	21	AAAG21039	Arabidopsis thaliana
19	81	8.2	346	22	ABAB71526	Drosophila melanog
20	81	8.2	486	21	AAAG21038	Arabidopsis thaliana
21	81	8.2	1137	14	AAAR42081	Imptiensis Necrotic
22	79	8.0	1290	22	ABBB62818	Drosophila melanog
23	76.5	7.7	433	13	AAAR62744	Alkali-protease Ya
24	76.5	7.7	433	19	AAAM61495	Modified Bacillus
25	76.5	7.7	433	20	AAAB95698	Bacillus sp. Lion
26	76.5	7.7	433	21	AAAY92207	Bacillus sp. Lion
27	76.5	7.7	433	21	AAAY46519	Amino acid sequenc
28	76.5	7.7	499	13	AAAR37574	Bacillus Lion Y en
29	76.5	7.7	636	26	AAAB99548	The AbF-B from A.
30	76.5	7.7	726	21	AAAB36463	Bacillus sp. alkali
31	76.5	7.7	849	17	AAAM6725	Human plakophilin
32	76	7.7	1237	18	AAAM55640	FLOI protein, invo
33	76	7.7	1237	20	AAAY17187	H. pylori ORF 04ep
34	75.5	7.6	244	16	AAAR70731	H. pylori Outer me
35	75.5	7.6	244	16	AAAR70732	Human protein S de
36	75.5	7.6	244	16	AAAR70733	Human protein S de
37	75.5	7.6	390	16	AAAR70730	Human protein S de
38	75.5	7.6	468	21	AAAB35882	Synthetic 55kd i-r
39	75.5	7.6	565	16	AAAR70729	Human protein S de
40	75.5	7.6	675	16	AAAR70728	Human protein S.
41	75.5	7.6	1219	22	AAAE03801	Human DNA helicase
42	74.5	7.5	763	22	ABBB62815	Drosophila melanog
43	74.5	7.5	1047	22	ABBB59194	Drosophila melanog
44	74.5	7.5	1247	22	ABBB61174	Drosophila melanog
45	74	7.5	211	21	AAAG25261	Arabidopsis thaliana

ALIGNMENTS

XX	AAAB84386
XX	ID AAAB84386 standard; Protein; 185 AA.
XX	AC AAAB84386;
XX	DT 22-AUG-2001 (first entry)
XX	DE Amino acid sequence of a protein-deamidating enzyme.
XX	KW Protein-deamidating enzyme; mineral absorption; food allergy; dough bakery; confectionery.
XX	OS Cryoseobacterium sp.
XX	PN EP1106696-A1.
XX	PD 13-JUN-2001.
XX	PF 04-DEC-2000; 2000EP-0310768.
XX	PR 03-DEC-1999; 99JP-0345044.
XX	PA (AMAN-) AMANO ENZYME INC.
XX	PI Yamaguchi S;
XX	DR WPI; 2001-376907/40.
XX	DR N-PSDB; AAP90280.
PT	New enzyme for use in e.g. bakery has an ability to deaminate amido groups in a protein -.
PS	Claim 4; Page 22; 43pp; English.

XX The present sequence represents a protein-deamidating enzyme from
 CC Cryoseobacterium sp. number 9670. The enzyme is able to deamidate amido
 CC groups in a protein by directly acting upon the amido groups without
 CC cutting peptide bonds and without cross-linking the protein. The enzyme
 CC thus reduces the mineral sensitivity of the protein and increases the
 CC soluble mineral content in the protein-mineral solution, improving the
 CC absorption of minerals in the human body. The enzyme reduces or removes
 CC the toxicity and allergenic property of the protein in a food
 CC e.g. allergy. The enzyme is useful for the improvement of dough in the
 CC field of bakery and confectionery e.g. for the production of crackers,
 CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean
 CC protein products, in various food articles e.g. meat or fish products
 CC and noodles; and for improving functionality of plant or animal protein.
 CC
 XX Sequence 185 AA;

Query Match 100.0%; Score 991; DB 22; Length 185;
 Best Local Similarity 100.0%; Pred. No. 2.3e-94;
 Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LASVIPDVATLNSLFNQKNSCGTSTASPCITFRYPVDCGYARAHKROIIMNNGYDC 60
 |||||||
 DB 1 lasvipdvatlnslfnqkngcgststaspcitfrypvdcgyarahkroqilimnngydc 60
 OY 61 EKQFYGNLKASTGTCCVAMSYHVALVSYKNASGVTEKRIIDPSLFSGGPYDTAWRNA 120
 |||||||
 DB 61 ekqfygnlkaastgtccvawsyhvalvsvyknasgvtekrliidpslfsqpydtawrna 120
 OY 121 CVNTSCGSASVSSYANTGNVYRSPNSLYDNNLINTNCVLTFTFSLSCGSPAPADY 180
 |||||||
 DB 121 cvntscgsasvssyantaqnyvyrspnslydnnlntncvltftfslsgcspapadv 180
 OY 181 SSCGF 185
 |||||
 DB 181 sscgf 185

RESULT 2
 AAB84387
 ID AAB84387 standard; Protein: 320 AA.

AC AAB84387;
 XX
 DT 22-AUG-2001 (first entry)

XX Amino acid sequence of a protein-deamidating enzyme.

KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
 KW bakery; confectionery.

XX Cryoseobacterium sp.

OS
 XX Key Location/Qualifiers
 FH 1..135
 FT Region
 FT /note= "prepro region"

XX EPI106696-A1.

XX 13-JUN-2001.

XX 04-DEC-2000; 2000EP-0310768.

XX 03-DEC-1999; 99JP-0345044.

XX (AMANO-) AMANO ENZYME INC.

XX Yamaguchi S;

XX WPI: 2001-376907/40.

XX N-PSDB: AAF90281.

PT New enzyme for use in e.g. bakery has an ability to deamidate amido
 PT groups in a protein
 XX
 XX Example 11; Page 24; 43pp; English.

CC The present sequence represents a protein-deamidating enzyme from
 CC Cryoseobacterium sp. number 9670. The enzyme is able to deamidate amido
 CC groups in a protein by directly acting upon the amido groups without
 CC cutting peptide bonds and without cross-linking the protein. The enzyme
 CC thus reduces the mineral sensitivity of the protein and increases the
 CC soluble mineral content in the protein-mineral solution, improving the
 CC absorption of minerals in the human body. The enzyme reduces or removes
 CC the toxicity and allergenic property of the protein in a food
 CC e.g. allergy. The enzyme is useful for the improvement of dough in the
 CC field of bakery and confectionery e.g. for the production of crackers,
 CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean
 CC protein products, in various food articles e.g. meat or fish products
 CC and noodles; and for improving functionality of plant or animal protein.
 CC
 XX Sequence 320 AA;

Query Match 100.0%; Score 991; DB 22; Length 320;
 Best Local Similarity 100.0%; Pred. No. 4.9e-94;
 Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LASVIPDVATLNSLFNQKNSCGTSTASPCITFRYPVDCGYARAHKROIIMNNGYDC 60
 |||||||
 DB 136 lasvipdvatlnslfnqkngcgststaspcitfrypvdcgyarahkroqilimnngydc 195
 OY 61 EKQFYGNLKASTGTCCVAMSYHVALVSYKNASGVTEKRIIDPSLFSGGPYDTAWRNA 120
 |||||||
 DB 136 ekqfygnlkaastgtccvawsyhvalvsvyknasgvtekrliidpslfsqpydtawrna 255
 OY 121 CVNTSCGSASVSSYANTGNVYRSPNSLYDNNLINTNCVLTFTFSLSCGSPAPADY 180
 |||||||
 DB 256 cvntscgsasvssyantaqnyvyrspnslydnnlntncvltftfslsgcspapadv 315
 OY 181 SSCGF 185
 |||||
 DB 316 sscgf 320

RESULT 3
 AAY44582
 ID AAY44582 standard; Protein: 185 AA.

AC AAY44582;

XX 04-APR-2000 (first entry)

XX Chryseobacterium gleum protein-deamidating enzyme.

DE
 XX Protein deamidating enzyme; soil bacterium; deamidation activity;
 KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
 KW protein engineering; surface hydrophobicity; toxicity; allergic;
 KW mineral sensitivity; calcium; absorption; mineral enhancing agent.

XX Chryseobacterium gleum 'JCM 2410'.

XX EP976829-A2.

XX 02-FEB-2000.

XX 04-JUN-1999; 99EP-0304367.

XX 04-JUN-1998; 98JP-0173940.

XX (AMANO) AMANO PHARM KK.

XX Yamaguchi S, Matsura A;

XX WPI: 2000-118552/11.

DR N-PSDB; AA249494.
 XX
 XX New enzyme for modifying and improving the function of proteins and/or
 PT peptides has deamidating activity without causing cross linking -
 XX
 XX Claim 5; Page 24; 57pp; English.
 PS
 XX The present amino acid sequence is the protein-deamidating enzyme,
 CC isolated from a new strain of soil bacterium, *Chryseobacterium gleum*
 CC JCM 2410. This enzyme exerts the deamidation activity by directly acting
 CC upon side chain amido groups in the protein in bonded state and
 CC releasing side chain carboxyl groups and ammonia. It can deaminate high
 CC molecular weight proteins, without cross linking and cleavage of peptide
 CC bonds, to improve protein function. This sequence is used for protein
 CC engineering, to cause an increase in surface hydrophobicity and improve
 CC the function of a plant or animal protein. It can also be used to remove
 CC or reduce toxicity or allergic property of proteins in food, decrease
 CC mineral sensitivity of protein, to allow greater absorption into the body
 CC and to solubilise calcium for use in drinks and mineral enhancing agents.
 CC
 XX Sequence 185 AA:
 SO
 Query Match 88.2%; Score 874; DB 21; Length 185;
 Best Local Similarity 87.4%; Pred. No. 2.9e-82;
 Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;
 QY 3 SVIPDVATLNSLFNQIKNSGCTSTASPCITFRYPVDCYARAHKRMQILMNGYDCEK 62
 DB 3 SVIPDVLTLNSLFNQIKNSGCTSTASPCITFRYPVDCYARAHKRMQILMNGYDCEK 62
 QY 63 QFYVGNLKASTGTCCVAMSVHVALIVSYKNASGVTEKRIIDPSLFSSGPTDTAMRNACV 122
 DB 63 qfyvgnlraatgtccvswyvhvllvsfknsagivckrlidpslfssgptdtsawraact 122
 QY 123 NTSCGSASVSSYANTAGNVYRRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDVSS 182
 DB 123 ntscgsasvssyantagnvyrrspgslllydnmyntncvlnlftsslsqspapdvss 182
 QY 183 CGF 185
 DB 183 cgf 185
 RESULT 4
 ID AAY44583 standard; Protein: 319 AA.
 XX
 AC AAY44583;
 XX
 DT 04-APR-2000 (first entry)
 XX
 DE *Chryseobacterium gleum* protein-deamidating prepro-enzyme.
 XX
 KW Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;
 KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
 KM protein engineering; surface hydrophobicity; toxicity; allergic;
 KM mineral sensitivity; calcium; absorption; mineral enhancing agent.
 XX
 OS *Chryseobacterium gleum* 'JCM 2410'.
 XX
 FH Key Location/Qualifiers
 FT 1..134
 FT Region /label= Prepro-region
 FT 1..21 /note= "Comprises the Pre and Pro region"
 FT Peptide /label= Signal peptide
 FT 22..134 /note= "Corresponds to the Pre region"
 FT Peptide /note= "Corresponds to the Pro region"
 FT 135..319 /label= Mature protein-deamidating-enzyme
 FT Protein /note= "Deamidates amido groups in a protein"

XX
 PN EP976829-A2.
 XX
 PD 02-FEB-2000.
 XX
 XX 04-JUN-1999; 99EP-0304367.
 XX
 PR 04-JUN-1998; 98JP-0173940.
 XX
 XX (AMANO) AMANO PHARM KK.
 PA
 PA Yamaguchi S, Matsuura A;
 PI
 DR WPI: 2000-118552/11.
 DR N-PSDB; AA249495.
 XX
 PT New enzyme for modifying and improving the function of proteins and/or
 PT peptides has deamidating activity without causing cross linking -
 XX
 PS Example 26; Page 26; 57pp; English.
 CC The present amino acid sequence is the protein-deamidating
 CC prepro-enzyme, isolated from a new strain of soil bacterium,
 CC *Chryseobacterium gleum* JCM 2410. This enzyme exerts the deamidation
 CC activity by directly acting upon side chain amido groups in the protein
 CC in bonded state and releasing side chain carboxyl groups and ammonia. It
 CC can deaminate high molecular weight proteins, without cross linking and
 CC cleavage of peptide bonds, to improve protein function. This sequence is
 CC used for protein engineering, to cause an increase in surface
 CC hydrophobicity and improve the function of a plant or animal protein. It
 CC can also be used to remove or reduce toxicity or allergic property of
 CC proteins in food, decrease mineral sensitivity of protein, to allow
 CC greater absorption into the body and to solubilise calcium for use in
 CC drinks and mineral enhancing agents.
 CC
 XX Sequence 319 AA:
 SO
 Query Match 88.2%; Score 874; DB 21; Length 319;
 Best Local Similarity 87.4%; Pred. No. 6e-82;
 Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;
 QY 3 SVIPDVATLNSLFNQIKNSGCTSTASPCITFRYPVDCYARAHKRMQILMNGYDCEK 62
 DB 137 SVIPDVLTLNSLFNQIKNSGCTSTASPCITFRYPVDCYARAHKRMQILMNGYDCEK 196
 QY 63 QFYVGNLKASTGTCCVAMSVHVALIVSYKNASGVTEKRIIDPSLFSSGPTDTAMRNACV 122
 DB 197 qfyvgnlraatgtccvswyvhvllvsfknsagivckrlidpslfssgptdtsawraact 256
 QY 123 NTSCGSASVSSYANTAGNVYRRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDVSS 182
 DB 257 ntscgsasvssyantagnvyrrspgslllydnmyntncvlnlftsslsqspapdvss 316
 QY 183 CGF 185
 DB 317 cgf 319
 RESULT 5
 ID AAB84385 standard; peptide: 20 AA.
 XX
 AC AAB84385;
 XX
 DT 22-AUG-2001 (first entry)
 XX
 DE Internal peptide of protein-deamidating enzyme.
 XX
 KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
 KW bakery; confectionery.
 XX
 OS *Cryseobacterium* sp.

```
XX  EP1106696-A1.
PN  13-JUN-2001.
XX
XX  04-DEC-2000; 2000EP-0310768.
XX  PE
XX  03-DEC-1999; 99JP-0345044.
XX  PR
XX  (AMANO-) AMANO ENZYME INC.
XX  PA
XX  Yamaguchi S;
XX  PI
XX  WPI: 2001-376907/40.
XX  DR
XX  New enzyme for use in e.g. bakery has an ability to deamidate amido
XX  PT groups in a protein -
XX
XX  Example 11: Page 20; 43pp; English.
XX  PS
XX  The present sequence represents a peptide fragment of a
XX  CC protein-deamidating enzyme from Cryoseobacterium sp. number 9670. The
XX  CC enzyme is able to deamidate amido groups in a protein by directly
XX  CC acting upon the amido groups without cutting peptide bonds and without
XX  CC cross-linking the protein. The enzyme thus reduces the mineral
XX  CC sensitivity of the protein and increases the soluble mineral content
XX  CC in the protein-mineral solution, improving the absorption of minerals
XX  CC in the human body. The enzyme reduces or removes the toxicity and
XX  CC allergenic property of the protein in a food e.g. allergy. The enzyme
XX  CC is useful for the improvement of dough in the field of bakery and
XX  CC confectionery e.g. for the production of crackers, biscuits, cookies,
XX  CC pizza pies or crusts of pie; in producing soybean protein products,
XX  CC in various food articles e.g. meat or fish products and noodles; and
XX  CC for improving functionality of plant or animal protein.
XX  SQ
XX  Sequence 20 AA:

Query Match 10.9%; Score 108; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  145 SPSNSYLDNNININCYLT 164
    |||||
DB  1 spnsylydnlnincvlt 20

RESULT 6
AAB84384
ID  AAB84384 standard; peptide: 20 AA.
XX
XX  AAB84384:
XX  AC
XX  22-AUG-2001 (first entry)
XX  DT
XX  N-terminal peptide of protein-deamidating enzyme.
XX  DE
XX  Protein-deamidating enzyme; mineral absorption; food allergy; dough;
XX  KW bakery; confectionery.
XX  KM
XX  Cryoseobacterium sp.
XX  OS
XX  EP1106696-A1.
XX  PN
XX  13-JUN-2001.
XX  PD
XX  04-DEC-2000; 2000EP-0310768.
XX  PE
XX  03-DEC-1999; 99JP-0345044.
XX  PR
XX  (AMANO-) AMANO ENZYME INC.
XX  PA
XX  Yamaguchi S;
XX  PI
```

```
XX  WPI: 2001-376907/40.
XX  DR
XX  New enzyme for use in e.g. bakery has an ability to deamidate amido
XX  PT groups in a protein -
XX
XX  Example 11: Page 20; 43pp; English.
XX  PS
XX  The present sequence represents a peptide fragment of a
XX  CC protein-deamidating enzyme from Cryoseobacterium sp. number 9670. The
XX  CC enzyme is able to deamidate amido groups in a protein by directly
XX  CC acting upon the amido groups without cutting peptide bonds and without
XX  CC cross-linking the protein. The enzyme thus reduces the mineral
XX  CC sensitivity of the protein and increases the soluble mineral content
XX  CC in the protein-mineral solution, improving the absorption of minerals
XX  CC in the human body. The enzyme reduces or removes the toxicity and
XX  CC allergenic property of the protein in a food e.g. allergy. The enzyme
XX  CC is useful for the improvement of dough in the field of bakery and
XX  CC confectionery e.g. for the production of crackers, biscuits, cookies,
XX  CC pizza pies or crusts of pie; in producing soybean protein products,
XX  CC in various food articles e.g. meat or fish products and noodles; and
XX  CC for improving functionality of plant or animal protein.
XX  SQ
XX  Sequence 20 AA:

Query Match 9.7%; Score 96; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 LASVIPDVATLNSLRNOKN 20
    |||||
DB  1 lasvipdvatlnslnglkn 20

RESULT 7
AG52451
ID  AG52451 standard; Protein: 594 AA.
XX
XX  AG52451:
XX  AC
XX  18-OCT-2000 (first entry)
XX  DT
XX  Arabidopsis thaliana protein fragment SEQ ID NO: 66675.
XX  DE
XX  Protein identification; signal transduction pathway; metabolic pathway;
XX  KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX  KW termination sequence.
XX  KM
XX  Arabidopsis thaliana.
XX  OS
XX  EP1033405-A2.
XX  PN
XX  06-SEP-2000.
XX  PD
XX  25-FEB-2000; 2000EP-0301439.
XX  PE
XX  25-FEB-1999; 99US-0121825.
XX  PR 05-MAR-1999; 99US-0123180.
XX  PR 09-MAR-1999; 99US-0123548.
XX  PR 23-MAR-1999; 99US-0125788.
XX  PR 25-MAR-1999; 99US-0126264.
XX  PR 29-MAR-1999; 99US-0126785.
XX  PR 01-APR-1999; 99US-0127462.
XX  PR 06-APR-1999; 99US-0128234.
XX  PR 08-APR-1999; 99US-0128714.
XX  PR 16-APR-1999; 99US-0129845.
XX  PR 19-APR-1999; 99US-0130077.
XX  PR 21-APR-1999; 99US-0130449.
XX  PR 23-APR-1999; 99US-0130510.
XX  PR 28-APR-1999; 99US-0130891.
XX  PR 30-APR-1999; 99US-0131449.
XX  PR 99US-0132048.
```


PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148568.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR	21-JUL-1999;	9905-0145086;
PR	21-JUL-1999;	9905-0145088;
PR	21-JUL-1999;	9905-0145089;
PR	22-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145089;
PR	22-JUL-1999;	9905-0145192;
PR	23-JUL-1999;	9905-0145145;
PR	23-JUL-1999;	9905-0145218;
PR	23-JUL-1999;	9905-0145224;
PR	23-JUL-1999;	9905-0145276;
PR	26-JUL-1999;	9905-01452913;
PR	27-JUL-1999;	9905-0145918;
PR	27-JUL-1999;	9905-0145919;
PR	27-JUL-1999;	9905-0145919;
PR	28-JUL-1999;	9905-0145951;
PR	02-AUG-1999;	9905-0145386;
PR	02-AUG-1999;	9905-0145388;
PR	02-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0147038;
PR	04-AUG-1999;	9905-0147204;
PR	04-AUG-1999;	9905-0147302;
PR	05-AUG-1999;	9905-0147192;
PR	05-AUG-1999;	9905-0147260;
PR	06-AUG-1999;	9905-0147316;
PR	06-AUG-1999;	9905-0147415;
PR	09-AUG-1999;	9905-0147493;
PR	09-AUG-1999;	9905-0147935;
PR	10-AUG-1999;	9905-0148171;
PR	11-AUG-1999;	9905-0148319;
PR	12-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148565;
PR	13-AUG-1999;	9905-0148684;
PR	16-AUG-1999;	9905-0149368;
PR	17-AUG-1999;	9905-0149475;
PR	18-AUG-1999;	9905-0149426;
PR	20-AUG-1999;	9905-0149722;
PR	20-AUG-1999;	9905-0149723;
PR	23-AUG-1999;	9905-0149902;
PR	23-AUG-1999;	9905-0149930;
PR	25-AUG-1999;	9905-0150566;
PR	26-AUG-1999;	9905-0150684;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151303;
PR	31-AUG-1999;	9905-0151548;
PR	01-SEP-1999;	9905-0152363;
PR	01-SEP-1999;	9905-0152363;
PR	10-SEP-1999;	9905-0153070;
PR	13-SEP-1999;	9905-0153758;
PR	15-SEP-1999;	9905-0154018;
PR	16-SEP-1999;	9905-0154039;
PR	20-SEP-1999;	9905-0154779;
PR	22-SEP-1999;	9905-0155139;
PR	23-SEP-1999;	9905-0155486;
PR	24-SEP-1999;	9905-0155569;
PR	28-SEP-1999;	9905-0156458;
PR	29-SEP-1999;	9905-0156596;
PR	04-OCT-1999;	9905-0157117;
PR	13-OCT-1999;	9905-0158293;
PR	13-OCT-1999;	9905-0159294;
PR	13-OCT-1999;	9905-0159295;
PR	14-OCT-1999;	9905-0159329;
PR	14-OCT-1999;	9905-0159330;
PR	14-OCT-1999;	9905-0159637;
PR	14-OCT-1999;	9905-0159638;
PR	18-OCT-1999;	9905-0159584;

```

Query Match          9.2%  Score 91;  DB 21;  Length 554;
Best Local Similarity 24.6%  Pred. No. 1.1;
Matches 31;  Conservative 14;  Mismatches 45;  Indels 36;  Gaps 5.

QY 25 TSTASSPCTTFRY---PVD--GCYARAHKMRQILMNNGY---DCEKOFVYGNLKAST-- 73
   | : : | | | | : : : : | : | | |
Db 442 tnvddnpstvtfnyfkhpydigrvcaearlrvskvscnrlfnyltgdcqkqvhhmlslsvya 501

QY 74 -----GTCVAAVSYHVALVS-----YKNASGVTERRIIDPSLFS 108
   | : | | | | | | | | | | | | | | |
Db 502 nmlrpkqndtksmagfcdltvctlvhbgcglvgkvvsprkrklygdrflrvldgstfd 561

QY 109 GGPVMD 114
   | : | : |
Db 562 espqtn 567

```

RESULT	9	
AAB30815		
ID	AAB30815	standard; Protein; 477 AA.
XX		
AC	AAB30815;	
XX		
DT	02-APR-2001	(first entry)
XX		
DE	Amino acid sequence of a prion-like amyloidogenic protein.	
XX		
KN	SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;	
KW	aggregation; fibril; phenotypic alteration; gene therapy;	
KW	disease resistance; plant pigmentation; prion disease.	
XX		
OS	Saccharomyces cerevisiae.	
XX		
PN	WO200075324-A2.	
XX		
PD	14-DEC-2000.	
XX		
PF	09-JUN-2000;	2000WO-US15876.
XX		
PR	09-JUN-1999;	99US-0138833.
XX		
PA	(ARCH-) ARCH DEV CORP.	
XX		
PI	Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheidel T;	
XX		
DR	WPI; 2001-061723/07.	
XX		
PT	New nucleic acid encoding chimeric proteins with self-assembly	
PT	properties, useful e.g. for diagnosis and treatment of prion diseases,	
PT	also related aggregates, fibrils and polymers -	
XX		
PS	Claim 11; Page 162-163; 188pp; English.	
XX		

PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	8.6%; Score 85; DB 21; Length 264;
Best Local Similarity	24.6%; Pred. No. 1.6;
Matches 31; Conservative	13; Mismatches 46; Indels 36; Gaps 5;

OY	25 TSAASSPCTIFRY--PYD--GCTARAHKMKQILMN-----NGYCEKOFVYNLKAFT-- 73
Db	112 ltvnddnpvsftfnyfkhpvdldqrcvealrlvskvtskrflnlytgcdkqnvhkmislvka 171
OY	74 -----GTCCVAMSYHVALIVS-----YKNASGVTEKRITIDPSIFS 108
Db	172 nlnlrpkqlndtkemaqckdlvtlwnhyhgglvgkvspnrkvlgydrllrvldgsfld 231
OY	109 SGPVTDT 114
Db	232 espqtn 237

RESULT 11
AAG25395
ID AAG25395 standard; Protein; 292 AA.
XX AC AAG25395;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 29442.
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999;	99US-0131449.
PR 30-APR-1999;	99US-0132048.
PR 30-APR-1999;	99US-0132407.
PR 04-MAY-1999;	99US-0132484.
PR 05-MAY-1999;	99US-0132485.
PR 06-MAY-1999;	99US-0132486.
PR 06-MAY-1999;	99US-0132487.
PR 07-MAY-1999;	99US-0132863.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155133.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.

PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.6%; Score 85; DB 21; Length 292;
Best Local Similarity 24.6%; Pred. NO. 1.8;

Matches 31; Conservative 13; Mismatches 46; Indels 36; Gaps 5;

QY 25 TSTASSPCITFRY---PVD--GCYARAHKMRQILMN---NGYDCKQFYVGNLKAFT-- 73
| : : | | | | | : : : : | : | : | |

Db 140 tnvddnpsvtfnfkhpdqrcveairlvskvtskrflnytgcdkgnvkmkmslsyka 199
| : : | | | | | : : : : | : | : | |

OY 74 -----GTCCVAMSYPVALIVS-----YKNASGVTEKRIIDPSLIFS 108
| : : | | | | | : : : : | : | : | |

Db 200 nlnlrpkqndtkmaqfcckdvtvltwihyhgclvgkvspnrivlgvdlrividgstfd 259
| : : | | | | | : : : : | : | : | |

OY 109 SGPVTD 114
| : : | | | | | : : : : | : | : | |

Db 260 espqtn 265
| : : | | | | | : : : : | : | : | |

RESULT 12

AAG25394
ID AAG25394 standard; Protein: 294 AA.

XX AAG25394;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 29441.

XX Protein identification: signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.

XX		Disclosure: SEQ ID NO 18633; 21pp + Sequence Listing; English.
PS		
CC	XX	The invention relates to an isolated nucleic acid detection reagent
CC	CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	CC	useful in developmental biology and in elucidating cell signaling and
CC	CC	cell-cell interactions in higher eukaryotes for the development of
CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	CC	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC	CC	sequences (AB101840-AB16175) and the encoded proteins
CC	CC	(ABB57737-ABB22072).
CC	CC	The sequence data for this patent did not form part of the printed
CC	CC	specification, but was obtained in electronic format directly from WIPO
CC	CC	at ftp.wipo.int/pub/published_pct_sequences.
XX		
SQ	Sequence	3726 AA;
OY	Query Match	8.5%; Score 84; DB 22; Length 3726;
Db	Best Local Similarity	25.3%; Pred. No. 70;
Matches	46; Conservative	21; Mismatches 73; Indels 42; Gaps 7;
OY	22 SCGTSTASSPCTTFYRPVDCGYARAHKMQILMNGYDCEK-----OFVYGNLKAST--	73
Db	178 ssgrrsgssgpdnsqassdga-----ssgisogkstkastsssklakttgya	225
OY	74 GTCCVAMSYHVALIVSYKNKASGVTEKRIID-----PSLFSSGPGVTDTAMRNACVNNSC	126
Db	226 gtcssa-----kskassgltseattsgisgaclkalfatpatstglacalvpg-	276
OY	127 GSASVSSYANTAGNVYRSPSNSLYDNNLITNCVLTFRSLLSGC-----SPPADPY	180
Db	277 gssgggtfipisaalliraknsnkkfknlrlargewmlpstsklkqlnspvydnpspppl	336
OY	181 SS 182	
Db	337 as 338	
RESULT 14		
AAY35054		
ID AAY35054	standard; Protein; 927 AA.	
XX AC	AAY35054;	
XX DT	13-SEP-1999 (first entry)	
XX DE	Chlamydia pneumoniae surface exposed polypeptide.	
XX KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;	
KW RW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;	
KW XX	vaccine; neutralising epitope.	
OS OS	Chlamydia pneumoniae.	
PX PN	WO9927105-A2.	
XX PD	03-JUN-1999.	
XX PF	20-NOV-1998; 98WO-IB01890.	
XX PR	04-NOV-1998; 98US-0107078.	
PR PR	21-NOV-1997; 97FR-0014673.	
XX PA	(GEST) GENSET.	
XX PI	Griifals R;	
XX DR	WPI, 1999-357842/30.	
PT PT	Genome sequence of Chlamydia pneumoniae	
XX PS	Page 942-944; Disclosure; 1912pp; English.	

[illegible]

XX Claim 7; page 60-62; 115pp; English.
 PS
 CC This polypeptide comprises the novel 98.4 kDa surface exposed
 CC protein Omp10 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC AX06822) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see AAM8417-28), and nucleic acid sequences encoding them (see
 CC AX06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.
 XX
 S0 Sequence 928 AA:

Query Match 8.3%; Score 82; DB 20; Length 928;
 Best Local Similarity 21.8%; Pred. No. 17; Mismatches 69; Indels 72; Gaps 13;
 Matches 49; Conservative 35; Mismatches 69; Indels 72; Gaps 13;
 OY 10 TLNSLFNOIKNOSCGTSTA-SSPCITFRYPVDGCFARAHKMRQILMN--NGYDCE----- 61
 Db 57 tlyntgdslnagspaltasc--fketgnlsfgghyqfllqndagancftnta 114
 OY 62 -----KQFVY-----GNLKASTGTCVAMSYHVALVSYKNASG----- 95
 Db 115 anklstfgsfylnltgtatctgtaik-stgacslqnsyfcyfgnfsndngalqgs 173
 OY 96 -----VTEKRITDPSLFSGPFDTAMRNACVNTSCGSASVS--SYANT 137
 Db 174 slslslnplltaknkacqk--ggalystgqilt-----lnltlnsastsentann 222
 OY 138 AGNVYRSPNSNYLDNMLIN-TNCVLTRFSLSG--CSPSPAP 178
 Db 223 ggalystgqilt-----lnltlnsastsentann 222
 OY 223 ggalystgqilt-----lnltlnsastsentann 222
 Db 265

Search completed: June 27, 2002, 21:47:22
 Job time: 5510 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 20:18:52 ; Search time 38.03 Seconds
(without alignments)
118.820 Million cell updates/sec

Title: US-09-727-769a-6

Perfect score: 991

Sequence: 1 LASVIPDVALTNSLFPQIKN.....FSLSGCSPSPAPDVSSCGF 185

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	874	88.2	185	4 US-09-324-910-6	Sequence 6, Appl1
2	874	88.2	319	4 US-09-324-910-12	Sequence 12, Appl1
3	83	8.4	20	4 US-09-324-910-2	Sequence 2, Appl1
4	79	8.0	20	4 US-09-324-910-1	Sequence 1, Appl1
5	77.5	7.8	666	4 US-08-982-785A-11	Sequence 11, Appl1
6	76.5	7.7	433	4 US-09-104-623A-4	Sequence 4, Appl1
7	76.5	7.7	499	2 US-07-952-853-6	Sequence 6, Appl1
8	76.5	7.7	499	2 US-08-914-848-6	Sequence 6, Appl1
9	76.5	7.7	635	2 US-08-873-479-43	Sequence 43, Appl1
10	75	7.6	271	1 US-08-467-155A-10	Sequence 10, Appl1
11	75	7.6	271	2 US-08-628-198-10	Sequence 10, Appl1
12	75	7.6	271	4 US-09-201-038-10	Sequence 10, Appl1
13	75	7.6	271	5 PCT-US96-07343-10	Sequence 10, Appl1
14	74.5	7.5	440	1 US-07-930-686-12	Sequence 12, Appl1
15	74.5	7.5	440	2 US-08-460-998-12	Sequence 12, Appl1
16	73	7.4	659	4 US-08-894-818B-1	Sequence 1, Appl1
17	73	7.4	659	4 US-08-894-818B-5	Sequence 5, Appl1
18	73	7.4	659	4 US-09-445-472-12	Sequence 12, Appl1
19	72	7.3	615	2 US-08-525-742-10	Sequence 10, Appl1
20	72	7.3	862	1 US-08-325-267A-4	Sequence 4, Appl1
21	72	7.3	894	3 US-08-362-525-22	Sequence 22, Appl1
22	72	7.3	894	3 US-08-971-692-15	Sequence 15, Appl1
23	72	7.3	1537	1 US-08-325-267A-2	Sequence 2, Appl1
24	71.5	7.2	720	3 US-09-257-799-48	Sequence 48, Appl1
25	71.5	7.2	720	3 US-08-920-919A-48	Sequence 48, Appl1
26	71	7.2	422	2 US-08-712-072C-2	Sequence 2, Appl1
27	69	7.0	1627	1 US-07-665-792E-9	Sequence 9, Appl1

28	68.5	6.9	294	4 US-09-188-930-294	Sequence 294, App
29	68.5	6.9	544	2 US-08-932-376A-4	Sequence 4, Appl1
30	68.5	6.9	572	2 US-08-932-376A-2	Sequence 2, Appl1
31	68.5	6.9	583	4 US-08-481-190-19	Sequence 19, Appl1
32	68.5	6.9	583	5 PCT-US93-00869-19	Sequence 19, Appl1
33	68	6.9	1003	1 US-08-571-758-4	Sequence 4, Appl1
34	68	6.9	1003	1 US-08-909-984A-4	Sequence 4, Appl1
35	68	6.9	1003	1 US-08-909-983-4	Sequence 4, Appl1
36	67.5	6.8	785	3 US-09-265-108-2	Sequence 2, Appl1
37	67.5	6.8	785	4 US-09-479-264-2	Sequence 2, Appl1
38	67	6.8	647	1 US-08-218-943-1	Sequence 1, Appl1
39	67	6.8	661	1 US-08-394-326-2	Sequence 2, Appl1
40	67	6.8	661	3 US-09-082-306-2	Sequence 2, Appl1
41	67	6.8	1041	1 US-08-220-151-4	Sequence 4, Appl1
42	67	6.8	1041	1 US-08-413-118-4	Sequence 4, Appl1
43	67	6.8	1041	3 US-08-473-446-4	Sequence 4, Appl1
44	67	6.8	1294	2 US-08-819-288-3	Sequence 3, Appl1
45	67	6.8	1294	4 US-09-400-348-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-09-324-910-6
: Sequence 6, Application US/09324910
: Patent No. 6251651
:
GENERAL INFORMATION:
: APPLICANT: Yamaguchi, Shotaro
: APPLICANT: Matsunura, Akira
: TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
: TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
: FILE REFERENCE: 0-54362
: CURRENT APPLICATION NUMBER: US/09/324,910
: CURRENT FILING DATE: 1999-06-03
: EARLIER APPLICATION NUMBER: HEI-10-173940
: EARLIER FILING DATE: 1998-06-04
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 6
: LENGTH: 185
: TYPE: PRT
: ORGANISM: Chryseobacterium gleum
:
US-09-324-910-6

Query Match      88.2%  Score 874;  DB 4;  Length 185;
Best Local Similarity 87.4%;  Pred. No. 1.5e-88;
Matches 160;  Conservative 10;  Mismatches 13;  Indels 0;  Gaps 0;

QY 3 SVIPDVALTNSLFPQIKNOSGCTSTPSPCTFRYPVDCYARAHKMKQILLNMGYDCK 62
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3 SVIPDLATNSLFPQIKNACGTSTPSPCTFRYPVDCYARAHKMKQILLNMGYDCK 62

QY 63 QFYVGNLKASTGTCVAVSHVAILVSYKNASGYTEKRIIDPSLSSGPTVDTAMRNACV 122
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 63 QFYVGNLRASGTCGVSVHVAILVSYFKNASGYTEKRIIDPSLSSGPTVDTAMRNACV 122

QY 123 NTSCGSAVSSTANTAGNAVYRSPNSLYLNNLINTNCVLTFRKSLSGCSPSPAPDVSS 182
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 123 NTSCGSAVSSTANTAGNAVYRSPNSGSLYDNNVYNTNCVINTFSLSGCSPPAPSVAS 182

QY 183 CGF 185
   |||
Db 183 CGF 185

RESULT 2
US-09-324-910-12
: Sequence 12, Application US/09324910
: Patent No. 6251651
:
GENERAL INFORMATION:
: APPLICANT: Yamaguchi, Shotaro
```

APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-324-910-12

Query Match 88.2%; Score 874; DB 4; Length 319;
Best Local Similarity 87.4%; Pred. No. 3.3e-88;

Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 3 SVIPVATLNSLFNQIKNSGCTSTASSPCITFRYPVDCGYARAHKMRQILANNGYDCER 62
DB 137 SVIPPLATLNSLFQIKNGACTSTASSPCITFRYPVDCGYARAHKMRQILNAGYDCER 196
QY 63 QFYVGNLKAAGTCCVANSYHVAIIIVSYKNASGVTEKRIIDPSLFSSGPTDTAMRACV 122
DB 197 QFYVGNLKAAGTCCVANSYHVAIIIVSYKNASGVTEKRIIDPSLFSSGPTDTAMRACV 256
QY 123 NMSCGSASVSANTAGNANYRSPSNLYDNNTLNNTNVLTKFKSLSCGSPSPAPDVSS 182
DB 257 NMSCGSASVSANTAGNANYRSPSNLYDNNTLNNTNVLTKFKSLSCGSPSPAPDVSS 316
QY 183 CGF 185
DB 317 CGF 319

RESULT 3
US-09-324-910-2
; Sequence 2, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-324-910-2

Query Match 8.4%; Score 83; DB 4; Length 20;
Best Local Similarity 78.9%; Pred. No. 0.0081;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 145 SPSNSLYDNNTLNNTNVL 163
DB 1 SPSGSLLYDNNTNNTNVL 19

RESULT 4
US-09-324-910-1
; Sequence 1, Application US/09324910

Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-324-910-1

Query Match 8.0%; Score 79; DB 4; Length 20;
Best Local Similarity 88.9%; Pred. No. 0.022;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 SVIPVATLNSLFNQIKN 20
DB 3 SVIPPLATLNSLFQIKN 20

RESULT 5
US-08-982-785a-11
; Sequence 11, Application US/08982785A
; Patent No. 6258929
; GENERAL INFORMATION:
; APPLICANT: Kosik, Kenneth S.
; APPLICANT: Zhou, Jianhua
; TITLE OF INVENTION: ALARM RELATED PEPTIDES AND
; TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,785A
; FILING DATE: 02-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,556
; FILING DATE: 02-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05311/018001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-982-785a-11

```

Query Match Similarity      7.8%; Score 77.5; DB 4; Length 666;
Best Local Similarity       27.5%; Pred No. 4.7;
Matches    38; Conservative   15; Mismatches     60; Indels    25; Gaps          6

OY      41 GCYA-RAHKROI LNNNGDCEKQFYGNLKA STGTCVAMSYHYVALIVSYKNASGVTEK 99
        || : : || ||| ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB      432 GCSNKSXDKM-----NMNNNDCLPBEELNPKGS-----GMLYHSDAIRLYLNLGMSKK 481

OY      100 RIIDSLEFSSGVPTDPAWMRNACVNYSGSAVSYS-YANTAGNVYYSPNSIYIDNNLIN 158
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      482 -----DATEACAGALONTTASKGSLSSGMSOLIGKEKGLPOLAIRLLSG--N 528

OY      159 TNCVLTKFLLSGCSPSP 176
        :: | : |||| | | |
DB      529 SDVVRSGASLLSNMCKP 546

RESULT      6
US-09-104-623A-4
: Sequence 4, Application US/09104623A
: Patent No. 6303752

GENERAL INFORMATION:
APPLICANT: Olsen, Arne Agerlin
APPLICANT: Fatum, Jine Muxoili
APPLICANT: Deussen, Heinz-Josef
APPLICANT: Roggen, Erwin Ludo
TITLE OF INVENTION: A Modified Polypeptide
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSSEE: NO. 63037520 NO. 6303752disk of NO. 6303752ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/104,623A
FILING DATE: 25-JUN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5256,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
STRAIN: Bacillus sp. Y
US-09-104-623A-4

Query Match      7.7%; Score 76.5; DB 4; Length 433;
Best Local Similarity 23.0%; Pred. No. 3.3;
Matches    45; Conservative   30; Mismatches     80; Indels    41; Gaps          7

OY      9 ATLNLSFNQIKNQSCGTASSPCITFRYPVDGCY-AFAHKMKROILNNGDCEKQFYVG 67
        |||:|||:| | | | | | | | | | | | | | | | | | | | | | | |
DB     108 SMLNTLFQAWMAAGARIHNIS-----WGAPRVNCAIYANRQVDVEYRRN--DWTYLFAG 160

OY      68 NIKASTGTCVAMSYHYVALIV-----SYK-----NASGVTERIIDPSLF 107

```

[illegible]

```

Db 166 -VLDQTHNDACCDYGAELFSSDTGTGHEHEATYLGNSITTWGAGADGPMIVMENN 224
Qy 93 ASGYTE-KRIIDPSL---FSSGCPITDPAIRACNTSCGSAVSXY----- 134
Db 225 FSGADEGINSDDPSLSTSYFYAAVAKGGADKAAIRGGNAASGLSTTYSGARPDYSGINPM 284
Qy 135 -----ANTAGNYITSPSNS-YLYDN--NLINTNCYLTKF---SLSGCS 173
Db 285 SKEGAILIGIGDMSNGAQGTFFYICGVMYTSGPSDDYENSVQENITVAARYGSLVSCPS 343

```

RESULT 8
US-08-914-848-6

```

1Sequence 6, Application US/08914848
2Patent No. 5989887
3
4GENERAL INFORMATION:
5APPLICANT: Van Heuvel, Margaretha
6APPLICANT: Bakhuis, Janna G.
7APPLICANT: Coutel, Yves
8APPLICANT: Harder, Abraham
9APPLICANT: De Graaff, Leendert H.
10APPLICANT: Filpphi, Michel J. A.
11APPLICANT: Van der Veen, Peter
12APPLICANT: Visser, Jacob
13APPLICANT: Andreoli, Peter M.
14TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
15TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
16TITLE OF INVENTION: ORIGIN
17NUMBER OF SEQUENCES: 24
18
19CORRESPONDENCE ADDRESS:
20ADDRESSEE: MORRISON & FOERSTER
21STREET: 755 Page Mill Road
22CITY: Palo Alto
23
24STATE: California
25ZIP: 94304-1018
26
27COMPUTER READABLE FORM:
28MEDIUM TYPE: Floppy disk
29COMPUTER: IBM PC compatible
30OPERATING SYSTEM: PC-DOS/MS-DOS
31SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
32CURRENT APPLICATION DATA:
33APPLICATION NUMBER: US/08/914,848
34

```

```

1      FILING DATE:
2      CLASSIFICATION:
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER:  US 07/952,853
5      FILING DATE: 25-NOV-1992
6      ATTORNEY/AGENT INFORMATION:
7      NAME:  MURSHIGE, Kate H.
8      REGISTRATION NUMBER:  29,959
9      REFERENCE/DOCKET NUMBER:  2461520035000
10     TELECOMMUNICATION INFORMATION:
11     TELEPHONE:  415-813-5600
12     TELEFAX:  415-494-0792
13     TELEX:  706141
14     INFORMATION FOR SEQ ID NO:  6:
15     SEQUENCE CHARACTERISTICS:
16     LENGTH: 499 amino acids
17     TYPE: amino acid
18     TOPOLOGY:  linear
19     MOLECULE TYPE:  protein
20     OS-08-914-848-6

```

Query match	7.78;	Score 76.5;	DB 2;	Length 499;
-------------	-------	-------------	-------	-------------

Matches 59; Conservative 22; Mismatches 77; Indels 81; Gaps 14;

QY 1 LASVIPDVATLN-----SLF-----NQIKNQSGTSTASSPCITFRYPVDGCYARAHKMR 50

Db 120 LASAIGAPVTLNGQKAYGVFMSPGTGYRNNEATGTATGDEP-----EGMYA----- 165

QY 51 QILMNGYDCEKQFVYCNLKAS--TGT-----CCVAMSYHVA----ILVSYKN-- 92

```

Db      :|||:|||||:|:|:|
166 -VLGSTRHNDACCEDYGAETSSDTTGAGHMEATLYLGNSTTWGCGAGDGRINAYDMENNL 224
Qy     93 ASGYTE-KRIIDPSL---FSSGPVTDIATRNACVNTSCGSASVS-----134
       |||||:|||||:|:|:|
Db     225 FSGADEGYNSDPSSISYFTVAAKVGADKMAIRGNAASGLSLTYSYGARPDYSGYNPM 284
Qy    135 -----ANTAGNYRSPSAS-LLYXN--NLINTNVLRKF---SYLSCS 173
       :|||:|||||:|:|:|
Db    285 SKEGAILIGICGDMSNGOGTFYEGCVMTSGPSPDDEIVSEVENTVAANKYVSGSLVLSGPS 343

```

RESULT 9
US-08-873-479-43

Sequence 43, Application US/08873479
 Patent No. 5891701
 GENERAL INFORMATION:
 APPLICANT: Sloima, Alan
 APPLICANT: Lynne, Christianson
 TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
 TITLE OF INVENTION: Having Protease Activity
 NUMBER OF SEQUENCES: 57
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/873,479
 FILING DATE: 12-JUN-1997
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Agt's, Cheryl H
 REGISTRATION NUMBER: 34,086
 REFERENCE/DOCKET NUMBER: 5251.000-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 43:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 635 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-873-479-43

Query Match	7.78; Score 76.5; DB 2; Length 635;
-------------	-------------------------------------

Matches 45; Conservative 30; Mismatches 80; Indels 41; Gaps 7;

QY 9 ATLSLNFQIKNOSCGTSTASSPCITFRYPVDCY-ARAHKMRQILMNNGYDCEKEQFVYG 67

Db 310 SNLNTLESQAWNAGARIHTNS-----WGAPVNGAYTANSRQVDEYVRNN--DMTVLEAAG 3622

QY 68 NLKASTGTCVAMSYHVALV---SYK-----NASCVTEKRIIDPSLF 107

Db 363 NEGPNSGTISAPCTAKNAITVGATENYRPSFGSIADPNPHIAQFSSRGATRDRGRIKPDVT 422

QY 108 SSGPVTDTAMRNACVNTSCGSASVSSYAN-----TAGNVYYRSPSNSLYLDNNLI 157

Db 423 APGTFILSARSSLAPDSSFWMYNSKYAYMGTSMTPIVAGNV--AQLRHEFIKNRGI 479

QY 158 NTNCVLTKESSLGCS 173

Db 480 TPKPSLIKALINAGAT 495

RESULT 10

US-08-467-155A-10

; Sequence 10, Application US/08467155A

; Patent No. 5736377

; GENERAL INFORMATION:

; APPLICANT: Band, Viola

; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED

; MOLECULES AND METHODS

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,155A

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 00398/100001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 271 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-467-155A-10

Query Match 7.6%; Score 75; DB 1; Length 271;

Best Local Similarity 23.1%; Pred. No. 2.5;

Matches 33; Conservative 20; Mismatches 46; Indels 44; Gaps 7;

QY 36 RYPVDC--YARAHKMOILMNNGYD-CEKOFYVGNLAKSTGTC-----VAM 80

Db 18 RYPIIECKATSQPH--QVSLNSGTHFCGSLVNMVNSAHCYQSRVEVRLGEHNSST 74

QY 81 SYHVAIIYSYKNASGVTEKRIL-----DPSLFSSG--PYTDTAMR 118

Db 75 RYPIIIQVTEGSEGFSSSRVIRHPNYSYNINDIMLIKSRPATLSTRYPPIINTYVQ 134

QY 119 NACVNTSGSA---SVSSYANT 137

Db 135 PVALPTSCAPAGTMCIVSGMGT 157

RESULT 11

US-08-628-198-10

; Sequence 10, Application US/08628198

; Patent No. 5843694

; GENERAL INFORMATION:

; APPLICANT: Band, Viola

; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED

; MOLECULES AND METHODS

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/628,198

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/467,155

FILING DATE: 06-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00398/100002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 271 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-628-198-10

Query Match 7.6%; Score 75; DB 2; Length 271;

Best Local Similarity 23.1%; Pred. No. 2.5;

Matches 33; Conservative 20; Mismatches 46; Indels 44; Gaps 7;

QY 36 RYPVDC--YARAHKMOILMNNGYD-CEKOFYVGNLAKSTGTC-----VAM 80

Db 18 RYPIIECKATSQPH--QVSLNSGTHFCGSLVNMVNSAHCYQSRVEVRLGEHNSST 74

QY 81 SYHVAIIYSYKNASGVTEKRIL-----DPSLFSSG--PYTDTAMR 118

Db 75 RYPIIIQVTEGSEGFSSSRVIRHPNYSYNINDIMLIKSRPATLSTRYPPIINTYVQ 134

QY 119 NACVNTSGSA---SVSSYANT 137

Db 135 PVALPTSCAPAGTMCIVSGMGT 157

RESULT 12

US-09-201-038-10

; Sequence 10, Application US/09201038

; Patent No. 6153387

; GENERAL INFORMATION:

; APPLICANT: Band, Viola

; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED

; MOLECULES AND METHODS

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-930-686-12

Query Match 7.5%; Score 74.5; DB 1; Length 440;
Best Local Similarity 23.1%; Pred. No. 5.6;
Matches 39; Conservative 22; Mismatches 59; Indels 49; Gaps 10;

QY 41 GCYARAHKMQILMNNGYDCEKQFYVGNLKAFTGTCVA-----WSYHVAL--VS 89
| : : : : |||| | : : : :
Db 51 GFAPKAARMLKV---GYDCE---VEANTAAVAKCKFEHDPPEQRNRYWGQNLMLLGGTN 103
| : : : : |||| | : : : :
QY 90 YKNAS-----GVTEKRIIDPSLSS--GPVTDRAWRNA---CVNTSCG 127
| : : : : |||| | : : : :
Db 104 YSKTESAKLSVQAWYWEIKMGVDPDENILTMVEFDRGVGHYTOVAMOSSDKIGCAVEWCP 163
| : : : : |||| | : : : :
QY 128 SAS-VSSYANTAGVNYRSPNSLYIDNNLINTNCVLTKFSLSGCSPS 175
| : : : : |||| | : : : :
Db 164 TMTLVACEYNPAGN-----RINHYYID--IGDPCCTTDEDCQCTGCTCS 204
| : : : : |||| | : : : :

RESULT 15

US-08-460-998-12
; Sequence 12, Application US/08460998
; Patent No. 5942413

GENERAL INFORMATION:

APPLICANT: Sharp, Phillip J
APPLICANT: Magland, Barry M
APPLICANT: Cobon, Gary S
TITLE OF INVENTION: Nematode Vaccine
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley and Lardner
STREET: suite 500, 3000 K Street, NW
CITY: Washington
STATE: DC

COUNTRY: United States of America
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,998
FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/930,686
FILING DATE: 06-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PK4486
FILING DATE: 06-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU92/00040
FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: Bent, Stephen A

REGISTRATION NUMBER: 29,768

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-460-998-12

Query Match 7.5%; Score 74.5; DB 2; Length 440;
Best Local Similarity 23.1%; Pred. No. 5.6;
Matches 39; Conservative 22; Mismatches 59; Indels 49; Gaps 10;

QY 41 GCYARAHKMQILMNNGYDCEKQFYVGNLKAFTGTCVA-----WSYHVAL--VS 89
| : : : : |||| | : : : :
Db 51 GFAPKAARMLKV---GYDCE---VEANTAAVAKCKFEHDPPEQRNRYWGQNLMLLGGTN 103
| : : : : |||| | : : : :
QY 90 YKNAS-----GVTEKRIIDPSLSS--GPVTDRAWRNA---CVNTSCG 127
| : : : : |||| | : : : :
Db 104 YSKTESAKLSVQAWYWEIKMGVDPDENILTMVEFDRGVGHYTOVAMOSSDKIGCAVEWCP 163
| : : : : |||| | : : : :
QY 128 SAS-VSSYANTAGVNYRSPNSLYIDNNLINTNCVLTKFSLSGCSPS 175
| : : : : |||| | : : : :
Db 164 TMTLVACEYNPAGN-----RINHYYID--IGDPCCTTDEDCQCTGCTCS 204
| : : : : |||| | : : : :

Search completed: June 27, 2002, 21:48:14
Job time: 5362 sec

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-551 <STO>
A:Cross-references: GB:AE004753; GB:AE004091; NID:G9949433; PIDN:AMG06698.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA3310
C:Superfamily: Escherichia coli y1dB protein

Query Match	9.3%	Score 92;	DB 2;	Length 551;
Best Local Similarity	21.6%;	Pred. No. 2.1;		
Matches 42;	Conservative 24;	Mismatches 54;	Indels 74;	Gaps 9

```

0Y      15  ENQJKNOSCGSTV--SSPCITFFRYPVDC--YARAK-----MHOILMNNGYDC 60
      274  FTQVR--SCGSTVSVYPCMSQFPREDYSDKRAKTHEGLDILQIRAGVQVLENNSDC 331
0Y      61  EKEFVYGNLKASTC-----CVANSHYVAILVSKNAGVTEK 99
      332  K-----GTCARVNRNDIKPTQSPFCDDGKKCDELSILVGL--QETIDELQDD 376
Db      0Y      100  RII-----DPSLESSGPTTDAMNACVNTSCGSASVSYAANTRAGNYTSPSSNSYL 151
      377  AIIVLHSDGSHGPFYTRIPREMERFOPVCTRNLQDSCSEELVN-----V 422

```

```
QY 152 YDNNLINTNCVLTk 165
    || :: | ||
Db 423 YDNTILYTDHFLTk 436
```

RESULT 3

surface protein type 51B - Paramoecium tetraurella
C:Species: Paramoecium tetraurella
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C:Accession: S50820
R:Scott, J.; Lebeck, C.; Forney, J.
Nucleic Acids Res. 22, 5079-5084, 1994
A:title: Analysis of the micronuclear B type surface protein gene in Paramoecium tetraure
A:Reference number: S50820; MUID:9509630
A:Accession: S50820
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2395 <SC>
A:Cross-references: EMBL:007603; NID:9467226; PIDD:AA81947.1; PID:9467227
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C:Genetics:
A:Genetic code: SCS5
A:Initons: 472/3; 1310/3; 1821/3
A:Superfamily: G surface protein

Query Match	9.3%	Score 92	DB 1	Length 2395
Best Local Similarity	22.4%	Pred. No. 11		
Matches 57	Conservative 22	Mismatches 76	Indels 100	Gaps 12

QY	10	TLNLFQNNKSGCTASTASPCITFRYPD-----	GCYARAHKMQIILNNG	57
		: : : : : : :		
Db	487	TVNSTNAGCYDKTICENSLAOTIC-----	DKDLNKACIKWKCKCYKRECVLASSITTAH	539
QY	58	YDCEKQFVYGNLAKSTGTCV-----	AMSUYHAIVLYSKNAGV-----	96
		: : : : : : :		
Db	540	ADCGTYDV-GCGLSNTGTCGVLPLKKEALTIEACNIRLQVTSYKSKYPLCGMNGSSCI	598	
QY	97	-----TEKRIIDPSLFSFG-----	-----PVTDTA-----	116
		: : : : : : :		
Db	599	DKACSTAPKTTATTSDCGTAKSGGVANNPNVNGSIQGGDLPITCAARKSTENCCEITRTGF	658	
QY	117	-----WR-----NACVYTSGSGSAYSSYANTAGNYYRRSPNSYELDNNLINNCVLTFRFSLL	169	
		: : : : : : :		
Db	659	PTCLMNSATSNACSEKSCSTASVTT-----TTGFLVFESNTNCLAA-----LSNACIAAN--NNA	710	

QY	170	SGCSPSPAPDVSSCG	184
Db	711	DGCIPKP----	SSCG 721

RESULT 4

Adhesion of calyx edges protein ACE [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50765
R:Araki, T.; Nakatani-Goto, M.
Submitted to the EMBL Data Library, May 1999
A:Description: ADHESION OF CALYX EDGES (ACE), cDNA.
A:Reference number: Z25224
A:Accession: T50765

```

A:Status: preliminary; translated from GB/EMBL/DDbJ
A:Molecule type: mRNA
A:Residues: 1,594 <EAA>
A:Cross-references: EMBL:AB027458; PIDN:BAA77837.1
A:Experimental source: cultivar Columbia
A:Genetics:
A:Gene: ACE
A:Map position: 1

```

Query Match	9.28;	Score	91;	DB	2;	Length	594;
Best Local Similarity	24.68;	Pred. No.	2.7;				
Matches	31;	Conservative	14;	Mismatches	45;	Indels	36;
						Gaps	5;

QY 25 TSTAASSPCLTFRY----PVD--GCYARAHKMRQILMNNGY----DDEKQFYVGNLKLAST-- 73
| : : | | | : : : : : | : | : |
Db 442 TIVDDNPSTYFENYFKHVPVDLQRCVEAIRLVSQVTSNRLNLTQCDKQNHAKMLSLSYKA 501

QY 74 -----GTCCVANSYHVALVS-----YKNSGVTEKRIIDPSLFS 108

QY	109	SGPVTD	114
		:	
Db	562	ESPGTN	567

RESULT 5

hypothetical protein C53B7.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T28602

submitted to the EMBL Data Library, December 1995
A;Description: The sequence of C. elegans cosmid C53B7.

A;Reference number: Z20526

A:Accession: Y28802
A:Status: preliminary; translated from GR/EMBL/DBIT

```

!!status proteinarity)      0000000000 0000000000 0000000000 0000000000
A:Molecule type: DNA

```

A;Residues: 1-243 <DUZ>

A; Cross-references: EME
A; Experimental source:

```

C:\Genetics:
n:\bapgen\imc\cud\source\src\main\distcol\mz\cudone\cudone

```

A;Gene: CESP:C53B7.3

A;Map position: X
A;Introns: 8/3: 39/1: 134/1: 164/1

CONCLUSIONS:

Query match	9.18;	Score 90;	DB 2;	Length 243;
-------------	-------	-----------	-------	-------------

Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 43; Corresponding 14; Mismatches 64; Indels 40; Cons 64

QY 22 SCGTSTASSEPCITFRYPVDGCAVRAHKMRQILMNNGYDCEKQFVYGNIAKST--GTCCVA 79
| | : : : | | : : | | : | |
Db 39 SSGGTCSNGCCMSLSLNGGLXNTYN-----NNNOYDMNQ--YGNOMGTITGSGTYCT- 90

OY 80 MSYHVALVSYKNAAGVTEKRIIDPSLFS-----SGPV-----TDTAMRNACVN 123
 Db 91 -----TFSYSCRSCTGCTNNRNQSSYSNNMNMNSQYSNCGSTCTGCTYSDCYSGQCSN 142
 OY 124 TSCGSASVSYANTAGNVYRSPNSNY-LYDNNL-INTKCV 162
 Db 143 GVC-----VAQYGTSTNNAMYSSTSGNCTYDNCGLNORCV 179

RESULT 6

T21772

hypothetical protein T01D3.6 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T21772; T24296

R:Steward, C.

Submitted to the EMBL Data Library, November 1996

A:Reference number: Z19470

A:Accession: T21772

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-927 <WIL>

A:Cross-references: EMBL:Z81527; PIDN:CAB04279.1; GSPDB:GN00023; CESP:T01D3.6

A:Experimental source: clone F35E12

R:Steward, C.

Submitted to the EMBL Data Library, October 1996

A:Reference number: Z19870

A:Accession: T24296

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-927 <W12>

A:Cross-references: EMBL:Z81110; PIDN:CAB03262.1; GSPDB:GN00023; CESP:T01D3.6

A:Experimental source: clone T01D3

C:Genetics:

A:Gene: CESP:T01D3.6

A:Map position: 5

A:Introns: 38/2; 87/3; 128/2; 242/2; 374/3; 416/3; 734/3; 842/3; 857/1

Query Match

Best Local Similarity 9.0%; Score 89.5; DB 2; Length 927;

Matches 61; Conservative 25; Mismatches 68; Indels 149; Gaps 16;

OY 4 VIPDVATLNSLFNOI--KNQSCGTSTASPCITFRYPVDCYARAHKROI-----LMN 55
 Db 130 VCPDYASGRCCQNEIKKDKSCGNA-----DCYVANHQLNCTCRGYTARR 176
 OY 56 NGYDCEKQ-----FVYGNLKA--TGTCVAMS----- 81
 Db 177 NGRDCDKRVOQACMSGDPHVVTYDGLRFYDQCTPYVFSQPCTLPAPYLMYSVARAKNEL 236
 OY 82 -----YHNA-----ILVSYKN-----ASGVTEKRIIDPSLFS-----S 109
 Db 237 PGKGYHISQVSEVEVDLHNLTTHVDRSKTALVNGV--OVLTPWYPPNKNRTWTVRVRFS 293
 OY 110 GP-----VTDAMRNACV-----NTSCGSA-----SVSYANTAGNV 141
 Db 294 GFTFTIENDGQVVTFTTYSLSVCQVPDIPEFNGATTLCGLAGITDCKLDDVYVKNKGSV 353
 OY 142 Y-----YRSPSN-----SYLDNNLI-----NTNCVLTKFSLLS 170
 Db 354 LAIKSRQHPNNNNHADPMKTEDTWTIDKFLILRPGQENCINGQTLDMNTNCVSTISLAAQ 413
 OY 171 GCS 173
 Db 414 SCA 416

RESULT 7

T15881

hypothetical protein D1044.3 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15881

R:Pauley, A.

Submitted to the EMBL Data Library, June 1994

A:Description: The sequence of C. elegans cosmid D1044.

A:Reference number: Z18423

A:Accession: T15881

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1895 <PAU>

A:Cross-references: EMBL:U00065; NID:G495681; PID:G495684; PIDN:AAA50735.1; CESP:D104

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:D1044.3

A:Introns: 35/2; 63/2; 150/1; 191/3; 207/2; 405/3; 551/3; 629/1; 764/1; 933/1; 959/2;

Query Match

Best Local Similarity 9.0%; Score 89; DB 2; Length 1895;

Matches 46; Conservative 27; Mismatches 88; Indels 30; Gaps 9;

OY 20 NQSC-----GSTASPCITFRYPVDCYARAHKROI--LMNNGYDCEKQF--VY 66
 Db 1112 NQCVISNGLNCGITVSYNSQCITLASPQNCQTSQCIDNSYCMQMCCTCANNRYLVY 1171
 OY 67 GNKASTGTCCVAMSYHV---AIVSYKNAAGVTEKRIIDPSLFSGPVT---DTAMRN 119
 Db 1172 GYCVPIITSSICQOTQTLVNNQCVLLSTVGETCIANQCVGAGMKNSTCCCTNATAMYG 1231
 OY 120 ACVNTSCGSASVSYANTAGNVYR-----SPNSLYDNNLINTN--CYLTKFSLSGC 172
 Db 1232 YCISSSSSSCGN-SNQVISINGCMYTVQVGGSCFSQCLNNNAVCTNNICVSTFCSV--SC 1288
 OY 173 SPSPAPDVSSC 183
 Db 1289 STNQVCISNOC 1299

RESULT 8

G69212

conserved hypothetical protein MTH845 - *Methanobacterium thermoautotrophicum* (strainC:Species: *Methanobacterium thermoautotrophicum*

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: G69212

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.

J.; Liu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,

K.I.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7153, 1997

A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: fu

A:Reference number: A69000; MUID:98037514

A:Accession: G69212

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-283 <MTH>

A:Cross-references: GB:AE000861; GB:AE000666; NID:G2621930; PIDN:AA85343.1; PID:G262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH845

Query Match

Best Local Similarity 8.9%; Score 88.5; DB 2; Length 283;

Matches 29; Conservative 17; Mismatches 61; Indels 11; Gaps 2;

OY 41 GCYARAHKROIILMNNGYDCEKQFVYGNLKASTGT-----CCVAMSYHVALVSYKNA 93
 Db 161 GCNSTYDKAVRIFMNVBDSIDSYFYNTFRKGAVOTLHSGSANC-----DHTHLVALARA 216
 OY 94 SGVTEKRIIDPSLFSGPVTDAMRNACVNTSCGSASVSYANTAGNVYRSPNSYL 151
 Db 217 SGIPARIYMGNCYFRSGNTTIGHVWGQLYVNGRKYDALATSESNALGTVNMMDRSSAFI 274

RESULT 9

[illegible]

```

Db 157 PSITYNIPDLDAIYVVTASTIDKEFTKPLACVQVWLSNGRTVQTEYLSWNLVILTSGI 216
Oy 69 -----LKASTGTCCVAMSYHAILVSYKN-----ASGV 96
      | : | | : : : | : |
Db 217 MESVYYSLOGYTVTSTRLASYSISLLYEQNLAILAMISVSFLPPIVAAWTONFQMSMGI 276
      | : | : : | : |
Oy 97 TE---KRIIDPSLFSSGFTDTAMRNACV-----NTSCGSASVSYANTAGNVYRSP 146
      | : | : : | : |
Db 277 IRINFMRLEFDWYVATSGSPVYRNKEVLISYQKRSLNSKIISASSNLNG--IESSQ 334
      | : | : : | : |
Oy 147 SNSLYDNNLINTNCVLTFRSLSG 171
      | : | : : | : |
Db 335 KNDLITYSNLNSNDYLSKIIVLRG 359

```

Search completed: June 27, 2002, 21:49:25
 Job time: 5213 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:48:17 ; Search time 39.56 Seconds

(Without alignments)
181.070 Million cell updates/sec

Title: US-09-727-769a-6

Perfect score: 991

Sequence: 1 LASVIPDVATLNSLFNQIKN.....FSLSGCSPSPADVSSCGF 185

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	9.0	1895	YLK3_CAEEL	P19151 caenorhabdi
2	88.5	8.9	1251	YQ03_CAEEL	Q09550 caenorhabdi
3	86	8.7	344	CYS5_DICDI	P54640 dictyosteli
4	85.5	8.6	1082	SP23_YEAST	P35210 saccharomyc
5	82.5	8.3	5376	ZAN_MOUSE	O08709 mus musculu
6	82	8.3	928	PMP9_CHLUPN	Q92398 Chlamydia p
7	82	8.3	2715	G156_PARP	P13857 paramectum
8	79	8.0	221	GPI_ZINOF	P82473 zingiber of
9	77.5	7.8	422	SPM1_SCHPO	Q92398 schizosacch
10	77.5	7.8	728	PKP1_MOUSE	P97350 mus musculu
11	76.5	7.7	499	ABFP_ASPNG	P42255 aspergillus
12	76.5	7.7	505	IRK4_CAEEL	P52192 caenorhabdi
13	76.5	7.7	747	PKP1_HUMAN	Q13835 homo sapien
14	76	7.7	2704	G168_PARP	P17053 paramectum
15	75.5	7.6	603	CEAI_MOUSE	O61129 mus musculu
16	74.5	7.5	1046	PSTA_DICDI	P11976 dictyosteli
17	74.5	7.5	1047	HIRA_DROME	O17468 drosophila
18	74	7.5	364	YBBB_ECOLI	P33667 escherichia
19	74	7.5	1672	PMP9_CHLUPN	Q92398 Chlamydia m
20	74	7.5	3305	APLP_MANSE	Q25490 manduca sex
21	73.5	7.4	802	CSD2_ECOLI	P35313 escherichia
22	73.5	7.4	1679	YMF9_YEAST	O04958 saccharomyc
23	73	7.4	3726	TRX_DROME	P20659 drosophila
24	72.5	7.3	402	OPB4_CANAL	P46596 candida alb
25	72	7.3	313	YMS8_YEAST	Q03695 saccharomyc
26	72	7.3	533	CAG2_MOUSE	Q09200 mus musculu
27	72	7.3	730	GLN3_YEAST	P18494 saccharomyc
28	72	7.3	1110	YGLM_INSV	O01260 impatiens n
29	72	7.3	1136	C4BA_BACTI	P05519 bacillus th
30	72	7.3	1376	WGL2_CVMJC	P02385 murine coro
31	72	7.3	1537	FL01_YEAST	P32768 saccharomyc
32	72	7.3	4548	AP0A_HUMAN	P08519 homo sapien
33	71.5	7.2	666	YEAF_YEAST	P40002 saccharomyc

34	71.5	7.2	720	1	TGLX_HUMAN	O43548 homo sapien
35	71.5	7.2	1162	1	VGL2_IBVN	P11233 avian infec
36	71	7.2	217	1	CATS_BOVIN	P25336 bos taurus
37	71	7.2	406	1	LMP1_MOUSE	P14438 mus musculu
38	71	7.2	470	1	RAN1_SCHPO	P08092 schizosacch
39	71	7.2	741	1	HOX1_HALRO	P28468 halocynthia
40	70.5	7.1	201	1	YADL_ECOLI	P37017 escherichia
41	70.5	7.1	574	1	YB2A_SCHPO	P87311 schizosacch
42	70.5	7.1	690	1	TRFE_ORYLA	P79819 oryzias lat
43	70.5	7.1	725	1	NCA2_MOUSE	P13594 mus musculu
44	70.5	7.1	727	1	PKP1_BOVIN	Q28161 bos taurus
45	70.5	7.1	761	1	NCA2_HUMAN	P13592 homo sapien

ALIGNMENTS

RESULT	1	STANDARD	PRT	1895 AA.
YLK3_CAEEL	YLK3_CAEEL			
ID	YLK3_CAEEL			
AC	P19151			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Putative serine/threonine-protein kinase D1044.3 in chromosome III (EC 2.7.1.-).			
DE	(EC 2.7.1.-).			
GN	D1044.3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;			
OC	Rhabditidae; Pelodermidae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	Pauley A., Waterston R.;			
RL	Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
CC	STRONG, TO ZC84.1.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: U00065; AAA50735.1; -			
DR	HSSP: O63450; 1A06.			
DR	WormPep: D1044.3; CE01206.			
DR	InterPro: IPR002899; EB.			
DR	InterPro: IPR000719; Euk_Pkinase.			
DR	InterPro: IPR002290; Ser_thr_Pkinase.			
DR	Pfam: PF01683; EB; 12.			
DR	Pfam: PF00069; Pkinase; 1.			
DR	SMART: SM00220; S_TKc; 1.			
DR	SMART: SM00289; WR1; 12.			
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE: PS00108; PROTEIN_KINASE_ST; 1.			
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.			
KW	Hypothetical protein; Transferase; Serine/threonine-protein kinase;			
KW	ATP-binding.			
FT	DOMAIN 431	703		PROTEIN KINASE
FT	NP_BIND 437	445		ATP (BY SIMILARITY).
FT	BLINDING 461	461		ATP (BY SIMILARITY).
FT	ACT_SITE 569	569		BY SIMILARITY.
SQ	SEQUENCE 1895 AA; 208393 MW; F23C9F7881353AD6 CRC64;			

Query Match 9.0%; Score 89; DB 1; Length 1895;
Best Local Similarity 24.1%; Pred. No. 2.5;
Matches 46; Conservative 27; Mismatches 88; Indels 30; Gaps 9;

```

OY 20 NQSC-----GTSTASSPCITFRYPVDCIYARAHKM-ROILMNNNGYDCEKOF--VY 66
DB 1112 NQOCVSYNGLNQLCTGVSYNSOCITLSPGQNCOTSSQCIDNSVCMNCTNNNNRILYV 1171
OY 67 GNLKASTGTCVAMSVHY---AIIVSYKNASGVTEKRIIDPSLFSSGPT-----DTAMRN 119
DB 1172 GTCVPTTSSICQOTOTLVNQCVLISYGETCIANQOCVGMCSGTCOCCTNGATAMYG 1231
OY 120 ACVNTSCGSASVSYAMTANVYR-----SPSNSTYLDNNLINTN--CVLTKFSLTSGC 172
DB 1232 YCISSSSSSCN-SNOVSYNGMCTNYOVGGSCSFQOCLNANVCNNICVSTFCV--SC 1288
OY 173 SPSPADVSSC 183
DB 1289 STNOVCISMQC 1299

RESULT 2
YQ03_CAEEL STANDARD: PRT: 1251 AA.
ID YQ03_CAEEL
AC Q09550:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 133.5 kDa protein F26C11.3 in chromosome II.
GN F26C11.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Matthews P.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: Z47072; CA87369.1; -.
CC DR WormPep: F26C11.3; CE01561.
CC DR InterPro: IPR000436; Sush1_SCR_CCP.
CC DR Pfam: PF00084; Sush1.1.
CC DR SMART: SM00032; CCP.1.
CC DR Hypothetical protein.
CC FT DOMAIN 120 414
CC SEQUENCE 1251 AA; 133498 MW; 2B959ECA03B9954A CRC64;

Query Match 8.9%; Score 88.5; DB 1; Length 1251;
Best Local Similarity 23.7%; Pred. No. 1.7;
Matches 36; Conservative 24; Mismatches 55; Indels 37; Gaps 6;
OY 21 QSC-CTSTASSPCITFRYPVDCIYARAHKMROILMNNNGYDCEKOFVYGNLKASTGCVVA 79
DB 966 QGCKATSTTQTP-TTFNMPGTGT-TRILPSGRIILSES-----1021
OY 80 WSYHVAIIVSYKNASGVTEKRIIDPSLFSSGPTDTAMRNACVNTSCGSASVSYAMTANAG 139
DB 1022 -----LIAVKNCTVTLMQOLYNPSKMTTRET-TSDAEGCKANSSGTTSTMSGTGG 1073
OY 140 NVYRSPNSYLYDNNLINTNVCVLTFRFSLTSG 171
DB 1074 TTVSRKTTNSNPDIDSTLET-----TTFAMPGTG 1101

RESULT 3
CYS5_DICDI

```

```

ID CYS5_DICDI STANDARD: PRT: 344 AA.
AC P54640:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cysteine proteinase 5 precursor (EC 3.4.22.-).
GN CPRE OR CP5.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Ax4;
RX MEDLINE=96081966; PubMed=7499424;
RA Souza G.M., Hirai J., Mehta D.P., Freeze H.H.;
RT Identification of two novel Dictyostellium discoideum cysteine
RT proteinases that carry N-acetylglucosamine-1-P-modification."
RL J. Biol. Chem. 270:28938-28945(1995).
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN THE VEGETATIVE PHASE AND DECREASES
CC WITH THE START DEVELOPMENT, REAPPEARS IN LOW LEVELS WHEN THE
CC FRUITING BODY IS FORMED.
CC -1- PM: PHOSPHOGLYCOSYLATED, CONTAINS GLCNAC-ALPHA-1-P-SER RESIDUES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1. ALSO KNOWN AS THE
CC PAININ FAMILY OF THIOL PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: L36205; AAA92018.1; -.
CC DR HSRP: P07711; ICBL.
CC DR MEROPS: C01.081; -.
CC DR DictyDb: DD01061; CPRE.
CC DR InterPro: IPR000668; Peptidase_C1.
CC DR InterPro: IPR000169; Thiolprolact_site.
CC DR Pfam: PF00112; Peptidase_C1; 1.
CC DR PRINTS: PR00705; PAININ.
CC DR PROSITE: PS00139; THIOL_PROTEASE_CYS_1.
CC DR PROSITE: PS00639; THIOL_PROTEASE_HIS_1.
CC DR PROSITE: PS00640; THIOL_PROTEASE_ASN_1.
CC DR Hydrolase: Thiol proteinase; Lysosome; zymogen; glycoprotein;
CC KW Phosphorylation; Signal.
CC FT SIGNAL 1 17
CC FT PROPEP 18 111
CC FT CHAIN 112 344
CC FT DOMAIN 196 340
CC FT ACT_SITE 136 136
CC FT ACT_SITE 272 272
CC FT ACT_SITE 311 311
CC FT ACT_SITE 311 311
CC FT DISULFID 133 174
CC FT DISULFID 167 207
CC FT DISULFID 265 333
CC FT CARBOHD 110 110
CC FT CARBOHD 297 297
CC SEQUENCE 344 AA; 37212 MW; 82F3F5B8C147BA8 CRC64;

Query Match 8.7%; Score 86; DB 1; Length 344;
Best Local Similarity 25.4%; Pred. No. 0.66;
Matches 49; Conservative 19; Mismatches 73; Indels 52; Gaps 9;
OY 26 STASSPCITFRYPVDCIYARAHKMROILMNNNGYDCEKOFVYGNLKASTGTCVAMSY 82
DB 168 STENSGC-----DGLMTYA-----FEYIINNNSIDTSSYPY---KAENCKEYKSEN 213
OY 83 HVAIIVSYKNASGVTEKRI-----IDPS-----LFSSGPTDTAMRNAC----- 121
DB 214 SGATLSYKTYAGSESSLESAAVNVNVPVSAIDASHOSFOLYTSG-----IYIEPCSSSEN 269

```

Qy 122 -----VNTSCGASVSSYANTAGNVYRSPNSLYDNINLINTNCVLTFRSLSGC 172
 Db 270 LDHCVLAVGVGGSSGSSGSSGSSGSSGNL-SASSNNEYWIKYKNSWGTSMGTEGIIILSRN 328
 Qy 173 SPSPADPVSSCGF 185
 Db 329 RDNNCGIASSASF 341

RESULT 4

SP23_YEAST STANDARD; PRT; 1082 AA.
 ID SP23_YEAST
 AC P35210;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE SPT23 protein.
 GN SPT23 OR YKL020C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rieger M.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 136-752 FROM N.A.
 RX MEDLINE=94262317; PubMed=8203154;
 RA Burkett T.J., Garfinkel D.J.;
 RT "Molecular characterization of the SPT23 gene: a dosage-dependent
 RT suppressor of Ty-induced promoter mutations from Saccharomyces
 RT cerevisiae.";
 RL Yeast 10:81-92(1994).
 CC -1- FUNCTION: DOSAGE-DEPENDENT SUPPRESSOR OF TY-INDUCED PROMOTER
 CC MUTATIONS. MAY EXERT ITS SUPPRESSION EFFECT THROUGH PROTEIN-
 CC PROTEIN INTERACTIONS SINCE DOES NOT PRESENT ANY OF THE MOTIFS
 CC GENERALLY FOUND IN TRANSCRIPTIONAL ACTIVATORS OR DNA BINDING
 CC PROTEINS.
 CC -1- SIMILARITY: CONTAINS 2 ANK REPEATS.
 CC -1- SIMILARITY: TO YEAST MG2.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 738
 CC ONWARD AND IS SHORTER (752 AA) DUE TO A FRAMESHIFT.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: Z28020; CAAB1855.1; -;
 DR EMBL: L24760; AAA20575.1; ALT_FRAME.
 DR PIR: S37837; S37837.
 DR SGD: S0001503; SPT23.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR002909; IPT_TIG.
 DR Pfam: PF00023; ank; 2.
 DR Pfam: PF01833; TIG; 1.
 DR SMART: SM00248; ANK; 2.
 DR SMART: SM00429; IPT; 1.
 DR PROSITE: PSS0068; ANK_REPEAT; 2.
 DR PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
 DR ANK repeat; Repeat.
 KW ANK repeat; Repeat.
 FT REPEAT 709 738 ANK 1.
 FT REPEAT 742 771 ANK 2.
 FT CONFLICT 715 715 H -> P (IN REF. 2).
 FT SEQUENCE 1082 AA; 121337 MW; 620C688ECC0ECBD7 CRC64;

Query Match

8.6%; Score 85.5; DB 1; length 1082;

Best Local Similarity 21.7%; Pred. No. 2.7;

Matches 40; Conservative 27; Mismatches 54; Indels 63; Gaps 8;

Qy 27 TASSPCITFRPVP-GCYARAHKRLQILMNGYCECFYGYNLKASTGCVAMSTHYA 85
 Db 259 SGNSNCINFDPTRIVCYCRHK-----ATNGF-----VV 288
 Qy 86 ILVSYKNAAGVTEKRIIDPSLF-----SSGPTFAMRNACVNTSCGASVSSYANTAGN 140
 Db 289 LFLRDHNGDILAKTTDPPIKIMDKKNSNTTPTSTMAOVSPM--TNDTRSTSSQSD 346
 Qy 141 VYRS-----PSNSLYDNINLINTNCVL-----TRFSLSGCSPSPAP 178
 Db 347 LNFSEFPPLPSNS---KNFYISTNCMLDSNCONNNNDNDKNNIKXTWTAMNNRRHPPSP 403
 Qy 179 DVSS 182
 Db 404 NSSS 407

RESULT 5

ZAN_MOUSE STANDARD; PRT; 5376 AA.
 ID ZAN_MOUSE
 AC 088799; 008647;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zonadhesin precursor.
 GN ZAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=98123114; PubMed=9452463;
 RA Gao Z., Garbers D.L.;
 RT "Species diversity in the structure of zonadhesin, a sperm-specific
 RT membrane protein containing multiple cell adhesion molecule-like
 RT domains.";
 RL J. Biol. Chem. 273:3415-3421(1998).
 RN [2]
 RP SEQUENCE OF 4864-5376 FROM N.A.
 RC TISSUE=Testis;

RX MEDLINE=97271566; PubMed=9126492;
 RA Gao Z., Harumi T., Garbers D.L.;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN)."
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD.
 CC -1- TISSUE SPECIFICITY: IN TESTIS: PRIMARILY IN HAPLOID SPERMATOS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOEA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE WMF DOMAIN 2 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUCC2).
 CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 25 WMF DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U97068; AAC26680.1; -

DR EMBL: U83190; AAC53125.1; -

DR MGD: MGI:106656; Zan.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR003645; FOLN.

DR InterPro: IPR000998; MAM.

DR InterPro: IPR002919; TIL.

DR InterPro: IPR003328; TILA.

DR InterPro: IPR001007; VMFC.

DR InterPro: IPR001846; Vwd.

DR Pfam: PF00629; MAM; 3.

DR Pfam: PF01826; TIL; 25.

DR Pfam: PF02345; Tila; 25.

DR Pfam: PF00094; vwd; 4.

DR SMART: SM00274; FOLN; 11.

DR SMART: SM00137; MAM; 2.

DR SMART: SM00214; VMC; 17.

DR SMART: SM00216; VMD; 4.

DR PROSITE: PS00022; EGF_1; 1.

DR PROSITE: PS01186; EGF_2; 18.

DR PROSITE: PS00740; MAM_1; FALSE_NEG.

DR PROSITE: PS0060; MAM_2; 3.

KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion; Repeat.

KW SIGNAL.

FT CHAIN 1 17

FT DOMAIN 18 5376

FT TRANSMEM 5311 5337

FT DOMAIN 5338 5376

FT DOMAIN 45 210

FT DOMAIN 215 374

FT DOMAIN 377 542

FT DOMAIN 547 1170

FT DOMAIN 1171 1280

FT DOMAIN 1281 1669

FT DOMAIN 1670 2056

FT DOMAIN 2057 2459

FT DOMAIN 2460 2579

FT DOMAIN 2580 2699

FT DOMAIN 2700 2819

FT DOMAIN 2820 2939

FT DOMAIN 2940 3059

FT DOMAIN 3060 3179

FT DOMAIN 3180 3299

FT DOMAIN 3300 3416

FT DOMAIN 3417 3536

FT DOMAIN 3537 3656

FT DOMAIN 3657 3776

FT DOMAIN 3777 3892

FT DOMAIN 3893 4029

FT DOMAIN 4029 4148

FT DOMAIN 4149 4263

FT DOMAIN 4264 4384

FT DOMAIN 4384 4503

FT DOMAIN 4504 4623

FT DOMAIN 4624 4743

FT DOMAIN 4744 4863

FT DOMAIN 4864 5000

FT DOMAIN 5001 5120

FT DOMAIN 5121 5240

FT DOMAIN 5241 5360

FT DOMAIN 5361 5480

FT DOMAIN 5481 5600

FT DOMAIN 5601 5720

FT DOMAIN 5721 5840

FT DOMAIN 5841 5960

FT DOMAIN 5961 6080

FT DOMAIN 6081 6200

FT DOMAIN 6201 6320

FT DOMAIN 6321 6440

FT DOMAIN 6441 6560

FT DOMAIN 6561 6680

FT DOMAIN 6681 6800

FT DOMAIN 6801 6920

FT DOMAIN 6921 7040

FT DOMAIN 7041 7160

FT DOMAIN 7161 7280

FT DOMAIN 7281 7400

FT DOMAIN 7401 7520

FT DOMAIN 7521 7640

FT DOMAIN 7641 7760

FT DOMAIN 7761 7880

FT DOMAIN 7881 8000

FT DOMAIN 8001 8120

FT DOMAIN 8121 8240

FT DOMAIN 8241 8360

FT DOMAIN 8361 8480

FT DOMAIN 8481 8600

FT DOMAIN 8601 8720

FT DOMAIN 8721 8840

FT DOMAIN 8841 8960

FT DOMAIN 8961 9080

FT DOMAIN 9081 9200

FT DOMAIN 9201 9320

FT DOMAIN 9321 9440

FT DOMAIN 9441 9560

FT DOMAIN 9561 9680

FT DOMAIN 9681 9800

FT DOMAIN 9801 9920

FT DOMAIN 9921 10040

FT DOMAIN 10041 10160

FT DOMAIN 10161 10280

FT DOMAIN 10281 10400

FT DOMAIN 10401 10520

FT DOMAIN 10521 10640

FT DOMAIN 10641 10760

FT DOMAIN 10761 10880

FT DOMAIN 10881 11000

FT DOMAIN 11001 11120

FT DOMAIN 11121 11240

FT DOMAIN 11241 11360

FT DOMAIN 11361 11480

FT DOMAIN 11481 11600

FT DOMAIN 11601 11720

FT DOMAIN 11721 11840

FT DOMAIN 11841 11960

FT DOMAIN 11961 12080

FT DOMAIN 12081 12200

FT DOMAIN 12201 12320

FT DOMAIN 12321 12440

FT DOMAIN 12441 12560

FT DOMAIN 12561 12680

FT DOMAIN 12681 12800

FT DOMAIN 12801 12920

FT DOMAIN 12921 13040

FT DOMAIN 13041 13160

FT DOMAIN 13161 13280

FT DOMAIN 13281 13400

FT DOMAIN 13401 13520

FT DOMAIN 13521 13640

FT DOMAIN 13641 13760

FT DOMAIN 13761 13880

FT DOMAIN 13881 14000

FT DOMAIN 14001 14120

FT DOMAIN 14121 14240

FT DOMAIN 14241 14360

FT DOMAIN 14361 14480

FT DOMAIN 14481 14600

FT DOMAIN 14601 14720

FT DOMAIN 14721 14840

FT DOMAIN 14841 14960

FT DOMAIN 14961 15080

FT DOMAIN 15081 15200

FT DOMAIN 15201 15320

FT DOMAIN 15321 15440

FT DOMAIN 15441 15560

FT DOMAIN 15561 15680

FT DOMAIN 15681 15800

FT DOMAIN 15801 15920

FT DOMAIN 15921 16040

FT DOMAIN 16041 16160

FT DOMAIN 16161 16280

FT DOMAIN 16281 16400

FT DOMAIN 16401 16520

FT DOMAIN 16521 16640

FT DOMAIN 16641 16760

FT DOMAIN 16761 16880

FT DOMAIN 16881 17000

FT DOMAIN 17001 17120

FT DOMAIN 17121 17240

FT DOMAIN 17241 17360

FT DOMAIN 17361 17480

FT DOMAIN 17481 17600

FT DOMAIN 17601 17720

FT DOMAIN 17721 17840

FT DOMAIN 17841 17960

FT DOMAIN 17961 18080

FT DOMAIN 18081 18200

FT DOMAIN 18201 18320

FT DOMAIN 18321 18440

FT DOMAIN 18441 18560

FT DOMAIN 18561 18680

FT DOMAIN 18681 18800

FT DOMAIN 18801 18920

FT DOMAIN 18921 19040

FT DOMAIN 19041 19160

FT DOMAIN 19161 19280

FT DOMAIN 19281 19400

FT DOMAIN 19401 19520

FT DOMAIN 19521 19640

FT DOMAIN 19641 19760

FT DOMAIN 19761 19880

FT DOMAIN 19881 20000

FT DOMAIN 20001 20120

FT DOMAIN 20121 20240

FT DOMAIN 20241 20360

FT DOMAIN 20361 20480

FT DOMAIN 20481 20600

FT DOMAIN 20601 20720

FT DOMAIN 20721 20840

FT DOMAIN 20841 20960

FT DOMAIN 20961 21080

FT DOMAIN 21081 21200

FT DOMAIN 21201 21320

FT DOMAIN 21321 21440

FT DOMAIN 21441 21560

FT DOMAIN 21561 21680

FT DOMAIN 21681 21800

FT DOMAIN 21801 21920

FT DOMAIN 21921 22040

FT DOMAIN 22041 22160

FT DOMAIN 22161 22280

FT DOMAIN 22281 22400

FT DOMAIN 22401 22520

FT DOMAIN 22521 22640

FT DOMAIN 22641 22760

FT DOMAIN 22761 22880

FT DOMAIN 22881 23000

FT DOMAIN 23001 23120

FT DOMAIN 23121 23240

FT DOMAIN 23241 23360

FT DOMAIN 23361 23480

FT DOMAIN 23481 23600

FT DOMAIN 23601 23720

FT DOMAIN 23721 23840

FT DOMAIN 23841 23960

FT DOMAIN 23961 24080

FT DOMAIN 24081 24200

FT DOMAIN 24201 24320

FT DOMAIN 24321 24440

FT DOMAIN 24441 24560

FT DOMAIN 24561 24680

FT DOMAIN 24681 24800

FT DOMAIN 24801 24920

FT DOMAIN 24921 25040

FT DOMAIN 25041 25160

FT DOMAIN 25161 25280

FT DOMAIN 25281 25400

FT DOMAIN 25401 25520

FT DOMAIN 25521 25640

FT DOMAIN 25641 25760

FT DOMAIN 25761 25880

FT DOMAIN 25881 26000

FT DOMAIN 26001 26120

FT DOMAIN 26121 26240

FT DOMAIN 26241 26360

FT DOMAIN 26361 26480

FT DOMAIN 26481 26600

FT DOMAIN 26601 26720

FT DOMAIN 26721 26840

FT DOMAIN 26841 26960

FT DOMAIN 26961 27080

FT DOMAIN 27081 27200

FT DOMAIN 27201 27320

FT DOMAIN 27321 27440

FT DOMAIN 27441 27560

FT DOMAIN 27561 27680

FT DOMAIN 27681 27800

FT DOMAIN 27801 27920

FT DOMAIN 27921 28040

FT DOMAIN 28041 28160

FT DOMAIN 28161 28280

FT DOMAIN 28281 28400

FT DOMAIN 28401 28520

FT DOMAIN 28521 28640

FT DOMAIN 28641 28760

FT DOMAIN 28761 28880

FT DOMAIN 28881 29000

FT DOMAIN 29001 29120

FT DOMAIN 29121 29240

FT DOMAIN 29241 29360

FT DOMAIN 29361 29480

FT DOMAIN 29481 29600

FT DOMAIN 29601 29720

FT DOMAIN 29721 29840

FT DOMAIN 29841 29960

FT DOMAIN 29961 30080

FT DOMAIN 30081 30200

FT DOMAIN 30201 30320

FT DOMAIN 30321 30440

FT DOMAIN 30441 30560

FT DOMAIN 30561 30680

FT DOMAIN 30681 30800

FT DOMAIN 30801 30920

FT DOMAIN 30921 31040

FT DOMAIN 31041 31160

FT DOMAIN 31161 31280

FT DOMAIN 31281 31400

FT DOMAIN 31401 31520

FT DOMAIN 31521 31640

FT DOMAIN 31641 31760

FT DOMAIN 31761 31880

FT DOMAIN 31881 32000

FT DOMAIN 32001 32120

FT DOMAIN 32121 32240

FT DOMAIN 32241 32360

FT DOMAIN 32361 32480

FT DOMAIN 32481 32600

FT DOMAIN 32601 32720

FT DOMAIN 32721 32840

FT DOMAIN 32841 32960

FT DOMAIN 32961 33080

FT DOMAIN 33081 33200

FT DOMAIN 33201 33320

FT DOMAIN 33321 33440

FT DOMAIN 33441 33560

FT DOMAIN 33561 33680

FT DOMAIN 33681 33800

FT DOMAIN 33801 33920

FT DOMAIN 33921 34040

FT DOMAIN 34041 34160

FT DOMAIN 34161 34280

FT DOMAIN 34281 34400

FT DOMAIN 34401 34520

FT DOMAIN 34521 34640

FT DOMAIN 34641 34760

FT DOMAIN 34761 34880

FT DOMAIN 34881 35000

FT DOMAIN 35001 35120

FT DOMAIN 35121 35240

FT DOMAIN 35241 35360

FT DOMAIN 35361 35480

FT DOMAIN 35481 35600

FT DOMAIN 35601 35720

FT DOMAIN 35721 35840

FT DOMAIN 35841 35960

FT DOMAIN 35961 36080

FT DOMAIN 36081 36200

FT DOMAIN 36201 36320

FT DOMAIN 36321 36440

FT DOMAIN 36441 36560

FT DOMAIN 36561 36680

FT DOMAIN 36681 36800

FT DOMAIN 36801 36920

FT DOMAIN 36921 37040

FT DOMAIN 37041 37160

FT DOMAIN 37161 37280

FT DOMAIN 37281 37400

FT DOMAIN 37401 37520

FT DOMAIN 37521 37640

FT DOMAIN 37641 37760

FT DOMAIN 37761 37880

FT DOMAIN 37881 38000

FT DOMAIN 38001 38120

FT DOMAIN 38121 38240

FT DOMAIN 38241 38360

FT DOMAIN 38361 38480

FT DOMAIN 38481 38600

FT DOMAIN 38601 38720

FT DOMAIN 38721 38840

FT DOMAIN 38841 38960

FT DOMAIN 38961 39080

FT DOMAIN 39081 39200

FT DOMAIN 39201 39320

FT DOMAIN 39321 39440

FT DOMAIN 39441 39560

FT DOMAIN 39561 39680

FT DOMAIN 39681 39800

FT DOMAIN 39801 39920

FT DOMAIN 39921 40040

FT DOMAIN 40041 40160

FT DOMAIN 40161 40280

FT DOMAIN 40281 40400

FT DOMAIN 40401 40520

FT DOMAIN 40521 40640

FT DOMAIN 40641 40760

FT DOMAIN 40761 40880

FT DOMAIN 40881 41000

FT DOMAIN 41001 41120

FT DOMAIN 41121 41240

FT DOMAIN 41241 41360

FT DOMAIN 41361 41480

FT DOMAIN 41481 41600

FT DOMAIN 41601 41720

FT DOMAIN 41721 41840

FT DOMAIN 41841 41960

FT DOMAIN 41961 42080

FT DOMAIN 42081 42200

FT DOMAIN 42201 42320

FT DOMAIN 42321 42440

FT DOMAIN 42441 42560

FT DOMAIN 42561 42680

FT DOMAIN 42681 42800

FT DOMAIN 42801 42920

FT DOMAIN 42921 43040

FT DOMAIN 43041 43160

FT DOMAIN 43161 43280

FT DOMAIN 43281 43400

FT DOMAIN 43401 43520

FT DOMAIN 43521 43640

FT DOMAIN 43641 43760

FT DOMAIN 43761 43880

FT DOMAIN 43881 44000

FT DOMAIN 44001 44120

FT DOMAIN 44121 44240

FT DOMAIN 44241 44360

FT DOMAIN 44361 44480

FT DOMAIN 44481 44600

FT DOMAIN 44601 44720

FT DOMAIN 44721 44840

FT DOMAIN 44841 44960

FT DOMAIN 44961 45080

FT DOMAIN 45081 45200

FT DOMAIN 45201 45320

FT DOMAIN 45321 45440

FT DOMAIN 45441 45560

FT DOMAIN 45561 45680

FT DOMAIN 45681 45800

FT DOMAIN 45801 45920

FT DOMAIN 45921 46040

FT DOMAIN 46041 46160

FT DOMAIN 46161 46280

FT DOMAIN 46281 46400

FT DOMAIN 46401 46520

FT DOMAIN 46521 46640

FT DOMAIN 46641 46760

FT DOMAIN 46761 46880

FT DOMAIN 46881 47000

FT DOMAIN 47001 47120

FT DOMAIN 47121 47240

FT DOMAIN 47241 47360

FT DOMAIN 47361 47480

FT DOMAIN 47481 47600

FT DOMAIN 47601 47720

FT DOMAIN 47721 47840

FT DOMAIN 47841 47960

FT DOMAIN 47961 48080

FT DOMAIN 48081 48200

FT DOMAIN 48201 48320

FT DOMAIN 48321 48440

FT DOMAIN 48441 48560

FT DOMAIN 48561 48680

FT DOMAIN 48681 48800

FT DOMAIN 48801 48920

FT DOMAIN 48921 49040

FT DOMAIN 49041 49160

FT DOMAIN 49161 49280

FT DOMAIN 49281 49400

FT DOMAIN 49401 49520

FT DOMAIN 49521 49640

FT DOMAIN 49641 49760

FT DOMAIN 49761 49880

FT DOMAIN 49881 50000

FT DOMAIN 50001 50120

FT DOMAIN 50121 50240

FT DOMAIN 50241 50360

FT DOMAIN 50361 50480

FT DOMAIN 50481 50600

FT DOMAIN 50601 50720

FT DOMAIN 50721 50840

FT DOMAIN 50841 50960

FT DOMAIN 50961 51080

FT DOMAIN 51081 51200

FT DOMAIN 51201 51320

FT DOMAIN 51321 51440

FT DOMAIN 51441 51560

FT DOMAIN 51561 51680

FT DOMAIN 51681 51800

FT DOMAIN 51801 51920

FT DOMAIN 51921 52040

FT DOMAIN 52041 52160

FT DOMAIN 52161 52280

FT DOMAIN 52281 52400

FT DOMAIN 52401 52520

FT DOMAIN 52521 52640

FT DOMAIN 52641 52760

FT DOMAIN 52761 52880

FT DOMAIN 52881 53000

FT DOMAIN 53001 53120

FT DOMAIN 53121 53240

FT DOMAIN 53241 53360

FT DOMAIN 53361 53480

FT DOMAIN 53481 53600

FT DOMAIN 53601 53720

FT DOMAIN 53721 53840

FT DOMAIN 53841 53960

FT DOMAIN 53961 54080

FT DOMAIN 54081 54200

FT DOMAIN 54201 54320

FT DOMAIN 54321 54440

FT DOMAIN 54441 54560

FT DOMAIN 54561 54680

FT DOMAIN 54681 54800

FT DOMAIN 54801 54920

FT DOMAIN 54921 55040

FT DOMAIN 55041 55160

FT DOMAIN 55161 55280

FT DOMAIN 55281 55400

FT DOMAIN 55401 55520

FT DOMAIN 55521 55640

FT DOMAIN 55641 55760

FT DOMAIN 55761 55880

FT DOMAIN 55881 56000

FT DOMAIN 56001 56120

FT DOMAIN 56121 56240

FT DOMAIN 56241 56360

FT DOMAIN 56361 56480

FT DOMAIN 56481 56600

FT DOMAIN 56601 56720

FT DOMAIN 56721 56840

FT DOMAIN 56841 56960

FT DOMAIN 56961 57080

FT DOMAIN 57081 57200

FT DOMAIN 57201 57320

FT DOMAIN 57321 57440

FT DOMAIN 57441 57560

FT DOMAIN 57561 57680

FT DOMAIN 57681 57800

FT DOMAIN 57801 57920

FT DOMAIN 57921 58040

FT DOMAIN 58041 58160

FT DOMAIN 58161 58280

FT DOMAIN 58281 58400

FT DOMAIN 58401 58520

FT DOMAIN 58521 58640

FT DOMAIN 58641 58760

FT DOMAIN 58761 58880

FT DOMAIN 58881 59000

FT DOMAIN 59001 59120

FT DOMAIN 59121 59240

FT DOMAIN 59241 59360

FT DOMAIN 59361 59480

FT DOMAIN 59481 59600

FT DOMAIN 59601 59720

FT DOMAIN 59721 59840

FT DOMAIN 59841 59960

FT DOMAIN 59961 60080

FT DOMAIN 60081 60200

FT DOMAIN 60201 60320

FT DOMAIN 60321 60440

FT DOMAIN 60441 60560

FT DOMAIN 60561 60680

FT DOMAIN 60681 60800

FT DOMAIN 60801 60920

FT DOMAIN 60921 61040

FT DOMAIN 61041 61160

FT DOMAIN 61161 61280

FT DOMAIN 61281 61400

FT DOMAIN 61401 61520

FT DOMAIN 61521 61640

FT DOMAIN 61641 61760

FT DOMAIN 61761 61880

FT DOMAIN 61881 62000

FT DOMAIN 62001 62120

FT DOMAIN 62121 62240

FT DOMAIN 62241 62360

FT DOMAIN 62361 62480

FT DOMAIN 62481 62600

FT DOMAIN 62601 62720

FT DOMAIN 62721 62840

FT DOMAIN 62841 62960

FT DOMAIN 62961 63080

FT DOMAIN 63081 63200

FT DOMAIN 63201 63320

FT DOMAIN 63321 63440

FT DOMAIN 63441 63560

FT DOMAIN 63561 63680

FT DOMAIN 63681 63800

FT DOMAIN 63801 63920

FT DOMAIN 63921 64040

FT DOMAIN 64041 64160

FT DOMAIN 64161 64280

FT DOMAIN 64281 64400

FT DOMAIN 64401 64520

FT DOMAIN 64521 64640

FT DOMAIN 64641 64760

FT DOMAIN 64761 64880

FT DOMAIN 64881 65000

FT DOMAIN 65001 65120

FT DOMAIN 65121 65240

FT DOMAIN 65241 65360

FT DOMAIN 65361 65480

FT DOMAIN 65481 65600

FT DOMAIN 65601 65720

FT DOMAIN 65721 65840

FT DOMAIN 65841 65960

FT DOMAIN 65961 66080

FT DOMAIN 66081 66200

FT DOMAIN 66201 66320

FT DOMAIN 66321 66440

FT DOMAIN 66441 66560

FT DOMAIN 66561 66680

FT DOMAIN 66681 66800

FT DOMAIN 66801 66920

FT DOMAIN 66921 67040

FT DOMAIN 67041 67160

FT DOMAIN 67161 67280

FT DOMAIN 67281 67400

FT DOMAIN 67401 67520

FT DOMAIN 67521 67640

FT DOMAIN 67641 67760

FT DOMAIN 67761 67880

FT DOMAIN 67881 68000

FT DOMAIN 68001 68120

FT DOMAIN 68121 68240

FT DOMAIN 68241 68360

FT DOMAIN 68361 68480

FT DOMAIN 68481 68600

FT DOMAIN 68601 68720

FT DOMAIN 68721 68840

FT DOMAIN 68841 68960

FT DOMAIN 68961 69080

FT DOMAIN 69081 69200

FT DOMAIN 69201 69320

FT DOMAIN 69321 69440

FT DOMAIN 69441 69560

FT DOMAIN 69561 69680

FT DOMAIN 69681 69800

FT DOMAIN 69801 69920

FT DOMAIN 69921 70040

FT DOMAIN 70041 70160

FT DOMAIN 70161 70280

FT DOMAIN 70281 70400

FT DOMAIN 70401 70520

FT DOMAIN 70521 70640

FT DOMAIN 70641 70760

FT DOMAIN 70761 70880

FT DOMAIN 70881 71000

FT DOMAIN 71001 71120

FT DOMAIN 71121 71240

FT DOMAIN 71241 71360

FT DOMAIN 71361 71480

FT DOMAIN 71481 71600

FT DOMAIN 71601 7

RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;	
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";	
RL	Nat. genet. 21:385-389(1999).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-AR39;	
RX	MEDLINE=20150255; PubMed=10684935;	
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,	
RA	White O., Hickey E.K., Peterson J., Unterback T., BERRY K., Bass S.,	
RA	Linder K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,	
RA	Gwynn M., Nelson S.W., DeBooy R., Kolonay J., McClarty G., Salzberg S.L.,	
RA	Eisen J., Fraser C.M.;	
RT	"Genome sequences of Chlamydia trachomatis Morn and Chlamydia	
RT	pneumoniae AR39.";	
RL	Nucleic Acids Res. 28:1397-1406(2000).	
RN	[4]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-J138;	
RX	MEDLINE=20330349; PubMed=10871362;	
RA	Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,	
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;	
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138	
RT	from Japan and CwL029 from USA.";	
RL	Nucleic Acids Res. 28:2311-2314(2000).	
CC	-I- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)	
CC	(POTENTIAL).	
CC	-I- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.	
CC		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	at the European Bioinformatics Institute. There are no restrictions on ways	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).	
CC		
DR	EMBL; AJ133034; CAB37069.1; -	
DR	EMBL; AE001628; AAD18591.1; -	
DR	EMBL; AE002192; AAF8163.1; -	
DR	EMBL; AP002546; BAA98655.1; -	
DR	TIGR; CP0306; -	
DR	InterPro; IPR003368; DUF145.	
DR	InterPro; IPR003357; OMP.	
DR	Pfam; PF02415; DUF145; 1.	
DR	Pfam; PF02385; OMP; 1.	
KW	Outer membrane; Signal; Multigene family; Complete proteome.	
FT	SIGNAL	
FT	CHAIN	
FT	27	928
FT	SEQUENCE	928 AA; 98332 MW; 59910A8BF04F12219 CRC64; .
FT	1	26
FT	PROBABLE OUTER MEMBRANE PROTEIN PMP9.	
FT	POTENTIAL.	

```

Query Match          8.3%; Score 82; DB 1; Length 926;
Best Local Similarity 21.8%; Pred No. 4.9;
Matches    49; Conservative   35; Mismatches   69; Indels   72; Gaps   13.

QY      10 TLNLSLFNQIKNSCGCTSTA-SSPCITTFPRPYVGCCYARAHKHQIIMN--NGYDCE----- 61
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     57 TIYNLTGVSVLTMASPLATASC--FRETGNLSFGGAGCYOFLQNIDAGANCFTTNA 114

QY      62 -----KOPYV-----GMLKASTGTGCCVAWSHYHAILVSYKNAG----- 95
       | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
Db     115 ANKLSPSGFSYLSTIQTTNATTGGAIR-STGCACISGSNYCFFQGNSNDNGALQS 173

QY      96 -----VEEKRIIDSLEPSGGVTDTAMRNACVNINSCGASYS--SYANT 137
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     174 SISLSLNPNLTPAKRAKAIÖK---GGALYSTGEIT-----INNTLNASSESENTAANN 222

QY      138 AGNVYYRSPSNRGYLEYDNNLIN-TNCVLTRFKSLLSG--CSPPAP 178
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     223 GCATP--TEASSFISSNKKAISTFINNSVTAITSATGCAIIYCSTSIP 265

```

ID	G156_PARPR	STANDARD:	PRT; 2715 AA.
AC	P13837;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	01-AUG-1990 (Rel. 15, Last annotation update)		
DE	156G surface protein precursor.		
GN	156G.		
OS	Paramacium primaurelia.		
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculiida;		
CC	Paramacium.		
XX	NCBI_TaxID=5886;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-156;		
RX	MEDLINE=87060934; PubMed=3783679;		
RA	Pirat A., Katinka M., Caron F., Meyer E.;		
RT	"Nucleotide sequence of the Paramacium primaurelia G surface protein.		
RL	A huge protein with a highly periodic structure.";		
J.	J. Mol. Biol. 189:47-60(1986).		
CC	-1 FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION		
CC	ANTIGEN OF PARAMECIUM PRIMAURELIA.		
CC	-1 SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.		
CC	-1 MISCELLANEOUS: IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC		
CC	STRUCTURE WITH 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD		
CC	CONTAINING 8 CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE		
CC	PART OF 475 RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE		
CC	MIDDLE OF THE PROTEIN.		
CC	-1 MISCELLANEOUS: EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES		
CC	(14-33 DEGREES CELSIUS).		
CC	-1 SIMILARITY: 98% TO THE ALLELIC FORM 166G PROTEIN (AC P17053) IN		
CC	PERIODIC STRUCTURE AND 80% IN VARIABLE DOMAIN IN THE MIDDLE OF		
CC	THE PROTEIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X03882; CAZ27514.1; .		
DR	PIR; A23475; A23475.		
DR	HSSP; P06620; INA.		
DR	InterPro; IPR002895; Paramacium_SA.		
DR	pfam; PF01508; Paramacium_SA_33.		
KW	Signal; Repeat; Antigen; Membrane; GPI-anchor.		
FT	SIGNAL	1	20
FT	POTENTIAL.		
FT	CHAIN	21	2715
FT	DOMAIN	106	2560
FT	SIMILAR	1	222
FT	PROTEIN		
FT	88% TO PARAMECIUM TERRAURELIA A		
SEQUENCE	2715 AA; 279551 MW; 97BE359AB9C7C298 CRC64;		

Query Match	8.3%	Score 82;	DB 1;	Length 2715;
Best Local Similarity	20.5%;	Pred. No. 17;		
Matches 51;	Conservative 23;	Mismatches 71;	Indels 104;	Gaps 11;

QY	11	LSNLFNQIKNSC-----GMSASPCTFRYPYVDG-----	41
	:	:::	:
	:	:::	:
Db	212	IQCWNKTKLTSCYWDGAACKDRIDCNAPTSLITTDACKTFR--IDGCTTKANGCGYTR	269
QY	42	--CYARAHKMQIILMNNGYD-----EKO-----FYGNLKASGT	75
	:	:	:
Db	270	TTCAARTIOASCIRKNSSGDCYMTGTACVDACANTPTTIAITNSACAGFVIGCITRSGGG	329
QY	76	CCVAMSYHVALLIVSKNNSGYTERIIPDSLFSSPVTDFAMRNACVMTSGSASV----	131
	:		:
Db	330	CVYNGACSYA-----NVOACVYKNPSNF-----DCIWDITCKEKEKCANASTTNN	374
QY	132	---SSVANT---AGNVYRSPSN-----SYLDNNILN-----TNCVL	163

DB 375 HDLCTSYSTCTVKGSGGCONFTCANADPTMTTNDACEAYFTGNNCTKRGSGGCVTNTTC 434
 QY 164 TKFSLSGC 172
 DB 435 AATLEAMC 443

RESULT 8

GPI_ZINOF
 ID GPI_ZINOF STANDARD: PRT: 221 AA.

AC P82473:
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Cysteine proteinase GP-I (EC 3.4.22.-).
 OS Zingiber officinale (Ginger).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;
 CC Zingiber.

NCBI_TaxID=94328;

RA MEDLINE=20156257; PubMed=10691991;

RA Choi K.H., Laursen R.A.;

RT "Amino-acid sequence and glycan structures of cysteine proteases with
 RT profile specifically from ginger rhizome zingiber officinale.";

RT Eur. J. Biochem. 267:1516-1526 (2000).

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE WITH A PROLINE RESIDUE
 AT P2.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
 PAPAIN FAMILY OF THIOL PROTEASES.

CC -1- CAUTION: THE AUTHORS REGARD THE SEQUENCE AS TENTATIVE, AS THEY
 BELIEVE THAT IT MAY HAVE BEEN CONTAMINATED BY A HOMOLOGOUS
 PROTEIN.

CC HSSP: P00785; 2ACT.

DR InterPro: IPR000668; Peptidase_C1.

DR InterPro: IPR000169; Thiolprot_act_site.

DR Pfam: PF00112; Peptidase_C1; 1.

DR PRINTS: PF00705; PAPAIN.

DR PROSITE: PS00139; THIOL_PROTEASE_CYS; FALSE_NEG.

DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.

DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.

KW HydroLase; Thiol protease; Glycoprotein.

FT ACT_SITE 27 27 BY SIMILARITY.

FT ACT_SITE 161 161 BY SIMILARITY.

FT DISULFID 24 65 BY SIMILARITY.

FT DISULFID 58 98 BY SIMILARITY.

FT DISULFID 155 206 BY SIMILARITY.

FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT UNSURE 2 2 OR D.

SO SEQUENCE 221 AA; 24241 MW; 3035D7870EA743DB CRC64;

Query Match 8.08; Score 79; DB 1; Length 221;
 Best Local Similarity 22.58; Pred. No. 1.8;
 Matches 36; Conservative 19; Mismatches 57; Indels 48; Gaps 8;

QY 18 IKNO-SCGTSTASSPCITFRFPVDS-----CYAAAH-----K 48
 DB 18 VKNQGGCGSCMAFDALIA-----VEGINQIVYGLDLSLEQQLVDCSTNNHGGGMPYRA 73
 QY 49 MROILMNGYDCEKOFYVGNLKAATGTCCVAMSYHVALIVSKNAGVTEKRIIDPSLFS 108
 DB 74 FOYLIINNGINSSEHYPR---TGITNGTCDTKENAHVVSIDSYRNPSPNDEKSL--QKAVA 128
 QY 109 SGAVTDT-----AMRNACVNTSCGSASVSYANTAG 139
 DB 129 NQPVSVTMDAAGRDFOLYRNGIFGTGSC--NISANHYRTVGG 167

RESULT 9
 SPML_SCHPO STANDARD: PRT: 422 AA.

AC 092398:
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Mitogen-activated protein kinase spml (EC 2.7.1.-) (MAP kinase spml)

DE (MAP kinase pmk1).

GN SPML OR PMK1 OR SPBC119.08.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

NCBI_TaxID=4896;

RA MEDLINE=97098653; PubMed=8943330;

RA Toda T., Dhut S., Superti-Furga G., Gotch G., Nishida E., Suglura R.,

RA Kuno T.;

RT "The fission yeast pmk1+ gene encodes a novel mitogen-activated
 RT protein kinase homolog which regulates cell integrity and functions
 RT coordinately with the protein kinase C pathway.";

RT Mol. Cell. Biol. 16:6752-6764 (1996).

RL [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=97280820; PubMed=9135147;

RA Zaitsevska-Carter T., Cooper J.A.;

RT "Spml, a stress-activated MAP kinase that regulates morphogenesis in
 RT S.pombe.";

RL EMBL J. 16:1318-1331 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Wood V., Rajandream M.A., Barrell B.G., Skellton J., Churcher C.M.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: REGULATES CELL INTEGRITY AND FUNCTIONS COORDINATELY WITH
 CC THE PROTEIN KINASE C PATHWAY (PKC1 AND PKC2). INVOLVED THE
 CC REGULATION OF WALL ARCHITECTURE, CELL SHAPE, CYTOKINESIS IN
 CC EXPONENTIAL AND STATIONARY PHASE, AND METABOLISM OF IONS.

CC -1- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE
 CC PHOSPHORYLATION BY SKH1/PEK1.

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: X98243; CAA66899.1; -

DR EMBL: U65405; AAC49707.1; -

DR EMBL: AL022117; CAAT1923.1; -

DR HSSP: Q16539; IMFC.

DR InterPro: IPR000719; Euk_kinase.

DR InterPro: IPR003527; MAP_kin.

DR InterPro: IPR002290; Ser_thr_pkinase.

DR Pfam: PF00069; pkinase; 1.

DR SMART: SM00220; S_TKC; 1.

DR PROSITE: PS01351; MAPK; 1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.

KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Cell cycle; Phosphorylation.

FT DOMAIN 21 314

FT NP_BIND 27 35

FT BINDING 52 52

FT ACT_SITE 149 149

FT MOD_RES 186 186

FT ACT_SITE 149 149 BY SIMILARITY.

FT MOD_RES 186 186 BY SIMILARITY.

FT MOD_RES 186 186 PHOSPHORYLATION (ACTIVATES THE KINASE)

FT (BY SIMILARITY).
 FT MOD_RES 188 188 PHOSPHORYLATION (ACTIVATES THE KINASE)
 FT DOMAIN 389 394 (BY SIMILARITY).
 FT SEQUENCE 422 AA; 48261 MW; FD02521E64E8BF82 CRC64;

Query Match 7.88; Score 77.5; DB 1; Length 422;
 Best Local Similarity 24.48; Pred. No. 5.2;
 Matches 29; Conservative 17; Mismatches 36; Indels 37; Gaps 6;

QY 84 VALIVSYKNSGVY---EKRIIDPSLF-----SSGPVTDAMRNAC 121
 DB 71 IKLIHFRNHNITCIVLDIINPYNENEYIYELEADLNLATIKSGQPLTDAHFQPSFI 130
 QY 122 VNTSCGSASVSYANTGNVYR--SPSNSTLYNNININCVL--TKFSLLSCGSPSP 176
 DB 131 YQILGGLKTHIS-----ANVIHRDLKPGNL-----LVNADCELKICDIFGLARGCSBNP 178

RESULT 10
 PKPI_MOUSE STANDARD; PRT; 728 AA.
 AC P97350;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Plakophilin 1.
 GN PKPI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RA [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=SKIN;
 RA Nimmitich V., Hunziker A.H., Franke W.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y07941; CAAG9240.1; -
 DR MGD: MGI:1328359; PKPI.
 DR Interpro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 3.
 DR PROSITE: PS50176; ARM_REPEAT; 3.
 KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
 KM Repeat.
 FT REPEAT 235 275 ARM 1.
 FT REPEAT 278 317 ARM 2.
 FT REPEAT 320 360 ARM 3.
 FT REPEAT 419 464 ARM 4.
 FT REPEAT 517 557 ARM 5.
 FT REPEAT 565 604 ARM 6.
 FT REPEAT 606 650 ARM 7.
 SQ SEQUENCE ~ 728 AA; 80896 MW; BDACSBA7B4118ACO CRC64;

Query Match 7.88; Score 77.5; DB 1; Length 728;
 Best Local Similarity 26.18; Pred. No. 9.8;

Matches 36; Conservative 18; Mismatches 59; Indels 25; Gaps 6;

QY 41 GCYA-BAKKKROLNNNGYCEKQFYVGNLKASGTCVAMSYVALVSYKNSGVYTER 99
 DB 492 GCFSNRGDKM-----MANNITDPLPEETNPKGS-----SWLHSDAIRTYLNMGSKK 541
 QY 100 RIIDPSLFSSGPVTDAMRNACVNTSCGSASVSYANTGNVYR--PSNSTLYNNIN 158
 DB 542 -----DNLKACAGALQNLTKSGKLSMGMSQLGKKEKGLPOLARLLQSG--N 588
 QY 159 TNCVLTNFSLLSCGSPSP 176
 DB 589 SDVVRGASGLSNMSRRP 606

RESULT 11
 ABFB_ASPNG STANDARD; PRT; 499 AA.
 AC P42255;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Alpha-L-arabinofuranosidase B precursor (EC 3.2.1.55) (Arabinosidase
 DE B) (ABF B).
 GN ABFB.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;

RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-30; 221-234; 285-304 AND
 RP 312-324.
 RC SPRAIN=CHS 120.49 / NA00;
 RA MEDLINE=94130338; PubMed=8299175;
 RA Filippi M.J.A., van Heuvel M., van der Veen P., Visser J.,
 RA de Graaf L.H.;
 RT "Cloning and characterization of the abfb gene coding for the major
 RT alpha-L-arabinofuranosidase (ABF B) of Aspergillus niger.";
 RL Curr. Genet. 24:525-532(1993).
 CC -1- FUNCTION: ABLE TO HYDROLYSE 1,5-, 1,3- AND 1,2-ALPHA-LINKAGES NOT
 CC ONLY IN L-ARABINOFURANOSYL OLIGOSACCHARIDES, BUT ALSO IN POLYSAC-
 CC CHARIDES CONTAINING TERMINAL NON-REDUCING L-ARABINOFURANOSSES IN
 CC SIDE CHAINS, LIKE L-ARABINAN, ARABINOGLACTAN AND ARABINOXYLAN.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
 CC arabinofuranoside residues in alpha-L-arabinosides.
 CC -1- PATHWAY: INVOLVED IN DEGRADATION OF THE PLANT CELL WALL
 CC POLYSACCHARIDE L-ARABINAN.
 CC -1- INDUCTION: BY GROWTH ON POLYMERIC SUBSTRATES AND L-ARABITOL.
 CC -1- SIMILARITY: BELONGS TO FAMILY 54 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X74777; CAAS2785.1; -
 DR EMBL: L23502; AAB53944.1; -
 DR EMBL: A27977; CAA01903.1; -
 KW Hydrolyase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 499 ALPHA-L-ARABINOFURANOSIDASE B.
 FT CARBOHYD 83 83 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 499 AA; 52523 MW; FIF6C3BA7AA41969 CRC64;

Query Match 7.78; Score 76.5; DB 1; Length 499;
 Best Local Similarity 24.78; Pred. No. 7.9;
 Matches 59; Conservative 22; Mismatches 77; Indels 81; Gaps 14;

```
OY 1 LASVDPVATLN-----SLF-----NOIKNOSCGTSTASSPCITFRYPVDCGYARAHKMR 50
DB 120 LASAIGAPVTLNGOKAYGVFMSPTGYRNNEATGTATGDEP-----EGMYA----- 165
OY 51 OILMANGYDCENGFYGNLKAS---TGT-----CCVAMSYHA---LLVSKN-- 92
DB 166 -VLDCGTHNDACCPCFYGNATISSTDTGAGHMEATILYNSTTWGAGDGPMIVDMENML 224
OY 93 ASGYTE-KRIIDPSL---FSSGPTVDTAMRNACVNTSCGSASVSSY----- 134
DB 225 FSGADEGVNSGSPSISYSVFTAAVAGADKMAIRGNMASSLSLTRYSGARDYSGYNPM 284
OY 135 -----ANTAGNYVRSPSNS-PLYDN--NLIINTCYLTKF---SLSGCS 173
DB 285 SKEGAIILIGIGDNGSNGAGTFEGVMTSGYPSDDVENSVCENIVAAKYSGSLVSGPS 343

RESULT 12
IRK4_CABEL STANDARD: PRT: 505 AA.
AC P52192;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inward rectifier potassium channel IRK-4.
GN IRK-4 OR R03E9.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilcox L.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: INWARD RECTIFIER K+ CHANNELS ARE CHARACTERIZED BY A
CC GREATER TENDENCY TO ALLOW POTASSIUM TO FLOW INTO THE CELL RATHER
CC THAN OUT OF IT.
CC -! SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -! SIMILARITY: BELONGS TO THE INWARD RECTIFIER-TYPE K+ CHANNEL
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: U40947; AAC48070.1; -.
DR WormPep: R03E9.4; CE04787.
DR InterPro: IPR001622; Channel_pore_K.
DR InterPro: IPR001838; KIR_channel.
DR Pfam: PF01007; IRK.1.
DR PRINTS: PR01320; KIRCHANNEL.
KW Ionc channel; Ion transport; Voltage-gated channel; Transmembrane;
KW Potassium transport.
FT DOMAIN 1 114 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 115 135 POTENTIAL.
FT DOMAIN 136 190 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 191 211 POTENTIAL.
FT DOMAIN 212 505 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 505 AA; 57419 MW; CIDF6DB2B280434 CRC64;

Query Match 7.7%; Score 76.5; DB 1; Length 505;
Best Local Similarity 22.1%; Pred. No. 8;
Matches 46; Conservative 27; Mismatches 78; Indels 57; Gaps 8;

OY 6 PDVATLNSLFNFIKNOGCGTSTASSPCITFRYPVDCGYARAHKMRQIILMNNYD----- 59
DB 285 PSIADDRLFLVWPTTLCHVIDSRSLVYN-----OQTLMSAQFETIIVLE 331
```

```
OY 60 --CENQFYGNLKAATGTCVAMSYHVALVSKNAGVTEKRIIDPSLF-SSGPTVDTA 116
DB 332 GIVESTGMTAQAKTSLVPEVLWGRFRKLVTRYQNSNGSYQ---IDYLFHSTYVPRPA 388
OY 117 W-----RNACVNTSCGSASVSSYANRAGNYVRSPSNS 149
DB 389 MSPAEFYSSKPNLKDYCHDSHEHKLIEDNRSSDSTPLSPSPYSPYPLN-HFOSSSNS 447
OY 150 VLYDNN--LINTNCYLTKFSLSGCSPS 175
DB 448 PVFSNNHKSFTNAVTCENGML--CPPT 473

RESULT 13
PKP1_HUMAN STANDARD: PRT: 747 AA.
AC Q13835; Q15152; Q00645;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Plakophilin 1 (Band-6-protein) (B6P).
GN PKP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-Epidermis;
RA MEDLINE=95074299; PubMed=7527055;
RA Hatfield M., Kristjansson G.I., Plessmann U., Weber K.;
RT "Band 6 protein, a major constituent of desmosomes from stratified
RT epithelia, is a novel member of the armadillo multigene family.";
RL J. Cell Sci. 107:2259-2270(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Bosch A.;
RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX PubMed=9369526;
RA Schmidt A., Langhein L., Rode M., Praetzel S., Zimbelmann R.,
RA Franke W.W.;
RT "Plakophilins 1a and 1b: widespread nuclear proteins recruited in
RT specific epithelial cells as desmosomal plaque components.";
RL Cell Tissue Res. 290:481-499(1997).
CC -! FUNCTION: SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES.
CC -! SUBCELLULAR LOCATION: Nuclear. Isoform 1 is also associated with
CC desmosomes.
CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/A AND 2/B (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -! TISSUE SPECIFICITY: NUCLEAR ISOFORM IS WIDELY EXPRESSED. ISOFORM 1
CC IS EXPRESSED IN STRATIFIED SQUAMOUS, COMPLEX, GLANDULAR DUCT AND
CC BLADDER EPITHELIA.
CC -! DISEASE: DEFECTS IN PKP1 ARE ASSOCIATED WITH ECTODERMAL
CC DYSPLASIA/SKIN FRAGILITY SYNDROME.
CC -! SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -! SIMILARITY: CONTAINS 8 ARM REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: X79293; CAA55881.1; -.
DR EMBL: Z34974; CAA84426.1; -.
DR EMBL: Z73677; -. NOT ANNOTATED_CDS.
DR EMBL: Z73678; CAA98022.1; -.
MIM: 601975; -.
```


DR MW: 604536; -
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 3.
 DR PROSITE: PSS0176; ARM_REPEAT; 3.
 KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
 FT Repeat: Alternative splicing.
 FT REPEAT 234 274 ARM 1.
 FT REPEAT 277 316 ARM 2.
 FT REPEAT 319 359 ARM 3.
 FT REPEAT 439 484 ARM 4.
 FT REPEAT 537 577 ARM 5.
 FT REPEAT 585 624 ARM 6.
 FT REPEAT 626 670 ARM 7.
 FT REPEAT 673 713 ARM 8.
 FT REPEAT 713 733 ARM 8.
 FT VARSPLIC 412 432 MISSING (IN ISOPFORM 1).
 FT CONFLICT 154 154 R -> G (IN REF. 1).
 FT CONFLICT 216 222 PPTSCNK -> RHLLQ (IN REF. 1).
 FT CONFLICT 462 462 V -> E (IN REF. 1).
 FT CONFLICT 496 496 Q -> K (IN REF. 1).
 FT CONFLICT 506 506 T -> P (IN REF. 1).
 FT CONFLICT 553 553 L -> S (IN REF. 1).
 SQ SEQUENCE 747 AA; 82860 MW; 60C1BCC50AB4E5F CRC64;

Query Match 7.7%; Score 76.5; DB 1; Length 747;
 Best Local Similarity 26.8%; Pred. No. 13;

Matches 37; Conservative 16; Mismatches 60; Indels 25; Gaps 6;

QY 41 GCVA-RAHKRQILMNGYDCERKQFYVGNLKAFTGTCVAMSYHVALVSKNAGYTEK 99
 DB 512 GCFNKSMDK-----MNNYDCPLPEEFNPKGS-----GWLHSDAIRYLNLMGSKK 561
 QY 100 RIIDPSLFSSCPYDTAMRNACVNTSCGSASVS-YANTANVYRSPNSLYLDNNLIN 158
 DB 562 -----DATEACAGALQNLTASKLSSGMSQLJLKEKGPOLARLLQSG--N 608
 QY 159 TNCVTRFSLSGCSPSP 176
 DB 609 SDVVRSGASLSLMSRHP 626
 RESULT 14
 G168_PARPR
 ID G168_PARPR STANDARD; PRT; 2704 AA.
 AC P17053;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-AUG-1990 (Rel. 15, Last annotation update)
 DE 168G surface protein precursor.
 GN 168G.
 OS Paramecium primaurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Paramecium.
 ON NCBI_TaxID=5886;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=90172419; PubMed=2308165;
 RA Prat A.;
 RT "Conserved sequences flank variable tandem repeats in two alleles of
 the G surface protein of Paramecium primaurelia.";
 RL J. Mol. Biol. 211:521-535(1990).
 CC -1- FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION
 ANTIGEN OF PARAMECIUM PRIMAURELIA.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- MISCELLANEOUS: IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC
 STRUCTURE WITH 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD
 CONTAINING 8 CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE
 PART OF 475 RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE
 MIDDLE OF THE PROTEIN.
 CC -1- MISCELLANEOUS: EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES
 (14-32 DEGREES CELSIUS).

CC -1- SIMILARITY: 98% TO THE ALLELIC FORM 156G PROTEIN (AC P13837) IN
 CC PERIODIC STRUCTURE AND 80% IN THE VARIABLE DOMAIN IN THE MIDDLE
 CC OF THE PROTEIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52133; CA36378.1; -
 DR PIR: S09118; S09118.
 DR HSSP: P06620; 11NA.
 DR InterPro: IPR002895; Paramecium_SA.
 DR Pfam: PF01508; Paramecium_SA; 33.
 KW Signal; Repeat; Antigen; Membrane; GPI-anchor.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2704 168G SURFACE PROTEIN.
 FT DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.
 FT DOMAIN 1060 1424 VARIABLE DOMAIN, COMPRISES 4 ALMOST
 FT IDENTICAL REPEATS.
 SQ SEQUENCE 2704 AA; 278775 MW; 40EA0A0B18EE2119 CRC64;

Query Match 7.7%; Score 76; DB 1; Length 2704;
 Best Local Similarity 19.9%; Pred. No. 61;

Matches 50; Conservative 22; Mismatches 71; Indels 108; Gaps 11;

QY 11 LNSLFOIKNOSC-----GSTASSPCITFRYPVDS----- 41
 DB 206 IQCVWNKILKTTSCYWDCAACKDRICDNAPTSLLTDDACKIFR--TDGCTTKANGCQYTR 263
 QY 42 --CYARAKRQILMNGYDC-----EKQ-----FVYGNLKASTGT 75
 DB 264 TTCAAATITQASCIKNSGCGYMTGACVDKTCANAPTTMTNSACAGFTGCTTKSGGG 323
 QY 76 CCVAMSYHVAL--VSKNAGYTEKRIIDPSLFSSCPYDTAMRNACVNTSCGSASV-- 131
 DB 324 CVANGACSVANVOAACVKNNSNF-----DCIWDTCRKRTCANAPTTN 366
 QY 132 -----SSYANT-----AGNYYRSPSN-----SYLDNNLIN-----TNC 161
 DB 367 NTHDLCTSYLSTCYVKGSGGCGQNRSCANAPTTMTNDACEAYLFGNCCITKSGGCVTNT 426
 QY 162 VLTKFSLSGC 172
 DB 427 TCAATITLCAAC 437
 RESULT 15
 CFAL_MOUSE
 ID CFAL_MOUSE STANDARD; PRT; 603 AA.
 AC 061129; O9MU07;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Complement factor I precursor (BC 3.4.21.45) (C3B/C4B inactivator).
 GN IF OR CFI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RX MEDLINE=96175003; PubMed=8604219;
 RA Minta J.O., Wong M.J., Kozak C.A., Kunath-Mugila L.M., Goldberger G.;
 RT "cDNA cloning, sequencing and chromosomal assignment of the gene for
 mouse complement factor I (C3b/C4b inactivator): identification of a
 RT species specific divergent segment in factor I.";
 RL Mol. Immunol. 33:101-112(1996).

100

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:47:27 ; Search time 102.74 Seconds
(without alignments)
311.505 Million cell updates/sec

Title: US-09-727-769a-6
Perfect score: 991
Sequence: 1 LASVLPDVATLNSLNFQIKN.....FSLSCGSPSPADPVSSCGF 185

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SPREMBL_19:*
- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	991	100.0	320	2	09A008
2	96.5	9.7	334	17	026459
3	92	9.3	551	16	09HYT4
4	92	9.3	2395	5	027167
5	91	9.2	594	10	09S746
6	90	9.1	243	5	018804
7	89.5	9.0	927	5	002364
8	89	9.0	672	5	0950P7
9	88.5	8.9	283	17	026933
10	85.5	8.6	534	10	09S3X3
11	85	8.6	594	10	09S3X3
12	84	8.5	342	13	P70004
13	84	8.5	3726	5	09VFL1
14	83.5	8.4	619	5	09GTE0
15	83.5	8.4	624	5	09GTD9
16	83.5	8.4	676	5	045184

17	83	8.4	95	5	09N992	09n992 leishmania
18	83	8.4	930	5	077151	077151 cryptospori
19	82.5	8.3	5374	11	099ND0	099nd0 mus musculu
20	82	8.3	703	3	008844	008844 saccharomyc
21	81.5	8.2	361	5	09XUH3	09xuh3 caenorhabdi
22	81.5	8.2	1360	12	09IKD1	09ikd1 rat sialoda
23	81	8.2	346	5	09VW94	09vw94 drosophila
24	81	8.2	485	10	09LX11	09lx11 arabisdopsis
25	80.5	8.1	508	10	09LTM8	09ltm8 arabidopsis
26	79.5	8.0	207	5	020608	020608 caenorhabdi
27	79.5	8.0	513	2	024719	024719 comamonas t
28	79.5	8.0	723	4	095789	095789 homo sapien
29	79.5	8.0	1316	17	09HLF6	09hlf6 thermoplasma
30	79.5	8.0	1894	11	P70206	P70206 mus musculu
31	79.5	8.0	1951	5	017042	017042 aequipecten
32	79	8.0	275	3	P78853	P78853 schizosacch
33	79	8.0	542	3	09P378	09p378 schizosacch
34	79	8.0	732	16	031447	031447 bacillus su
35	79	8.0	1290	5	09VTR8	09vtr8 drosophila
36	79	8.0	1314	5	09BLV9	09blv9 leishmania
37	78.5	7.9	370	3	094737	094737 pneumocysti
38	78.5	7.9	719	16	09CWN9	09cwn9 pasteurella
39	78.5	7.9	725	4	013835	013835 homo sapien
40	78.5	7.9	1067	5	016249	016249 caenorhabdi
41	77.5	7.8	400	5	0950P8	095qp8 caenorhabdi
42	77.5	7.8	539	12	091141	091141 avian infec
43	77.5	7.8	728	11	P97350	P97350 mus musculu
44	77.5	7.8	848	15	040366	040366 human immun
45	77.5	7.8	2717	5	094710	094710 paramecium

ALIGNMENTS

RESULT	ID	Query Match	Score	991: DB 2: Length	320: Best Local Similarity	100.0%: Pred. No. 1.8e-90; Mismatches 185; Conservative 0; Indels 0; Gaps 0;
1	09A008	100.0	320	2	09A008	
AC	09A008:	PRELIMINARY:	PRT:	320 AA.		
DT	01-JUN-2001 (TREMBLrel. 17, Created)					
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)					
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)					
DE	PROTEIN-GLUTAMINASE PRECURSOR.					
GN	PRGA.					
OS	Chryseobacterium proteolyticum.					
OC	Bacteria; CFB group; Flavobacteriia; Flavobacteriaceae;					
OC	Chryseobacterium.					
OX	NCBI_TaxID=118127;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=9670;					
RX	MEDLINE=21153247; PubMed=11231294;					
RA	Yamaguchi S., Jeenes D.J., Archer D.B.;					
RT	"Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme					
RT	that deamidates glutaminy residues in proteins: purification,					
RT	characterization and gene cloning.";					
RL	Eur. J. Biochem. 268:1410-1421(2001).					
DR	EMBL; AB046594; BAB21508.1; .					
KH	Signal.					
FT	SIGNAL. 1 21 POTENTIAL.					
FT	CHAIN 136 320					
SQ	SEQUENCE 320 AA; 35044 MW; C67823D2BC131410 CRC64;					

Qy	1	LASVLPDVATLNSLNFQIKNKGSTASTSPCTITFRYPVDCGYARAHKROILMNNNGDC	60
Db	136	LASVLPDVATLNSLNFQIKNKGSTASTSPCTITFRYPVDCGYARAHKROILMNNNGDC	195
Oy	61	EKQFYVGNLKASTGTCVAMSYHVALVSYKNASGVTEKRIIDPSLFSSGPTDTAWRNA	120

```

DB 196 ENQFYGNLKAATGTCVAMSHVALLVSKNASTGTEKRIIDPSLFSSGPTVDTAMRNA 255
QY 121 CVNTSCGASVSSSYANTAGNYYYRSPSNSLYLDNNLINTNCVLTFRSLSGCSPSPADPV 180
DB 256 CVNTSCGASVSSSYANTAGNYYYRSPSNSLYLDNNLINTNCVLTFRSLSGCSPSPADPV 315
QY 181 SSCGF 185
DB 316 SSCGF 320

RESULT 2
ID 026459 PRELIMINARY: PRT; 534 AA.
AC 026459;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
DE CONSERVED PROTEIN.
GN MTH359.
OS Methanothermobacter thermautotrophicus.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang N., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spatafore R., Viacre R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RA "Complete genome sequence of Methanothermobacter thermautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RT J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000821; AAB84865.1; -.
DR InterPro: IPR002931; Transglut_core.
DR Pfam: PF01841; Transglut_core; 1.
DR SMART; SM00460; Tgc; 1.
KW Complete proteome.
SQ SEQUENCE 534 AA; 57628 MW; 588412BDBD342E26 CRC64;

Query Match 9.7%; Score 96.5; DB 17; Length 534;
Best Local Similarity 24.1%; Pred. No. 0.24;
Matches 39; Conservative 23; Mismatches 71; Indels 29; Gaps 4;

QY 5 IDVATLNSLFNOIKNQ-----CGTSTASSPCITFRYPVD-----GCY 43
DB 358 LKNTYVTRFASLSLRPNDRPYRGESTARLYASSCPDSDIRSLASEITRGLTSTF 417
QY 44 ARAHK-----MROILMNGVDECKOFYGNLKAATGTCVAMSHVALLVSKNASTGTEK 99
DB 418 SRAEAVFGWNRNINYSFYNTKYGAVGTLKNRTGCV-----DTHLLVALRAAGIPAR 473
QY 100 RIIDPSLFSSGPTVDTAMRNACVNTSCGASVSSYANTAGN 141
DB 474 YHGTGNCFTSGNVYGHVMAQLLVGDTWYAADATSSRNSLGVV 515

RESULT 3
ID 09HYT4 PRELIMINARY: PRT; 551 AA.
AC 09HYT4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
DE HYPOTHETICAL PROTEIN PA3310.
GN PA3310.
OS Pseudomonas aeruginosa.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RA MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
DR EMBL; AE004753; AAG0698.1; -.
DR InterPro: IPR003371; DUF146.
DR Pfam; PF02418; DUF146; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 551 AA; 61421 MW; 3E3EAF5C308BE815 CRC64;

Query Match 9.3%; Score 92; DB 16; Length 551;
Best Local Similarity 21.6%; Pred. No. 0.71;
Matches 42; Conservative 24; Mismatches 54; Indels 74; Gaps 9;

QY 15 FNQIKNSCGTSTA-SSPCITFRYPVDCG---YARAK-----MROILMNGVDC 60
DB 274 FQYVR--SGGTSTAVSPCMFQYPRRSDSKAKTBELLDTLQKAGQVLMLENNSDC 331
QY 61 ENQFYGNLKAATGTC-----CVAMSHVALLVSKNASTGTEK 99
DB 332 K-----GTLRLVPRNDIPKTPSPDCGKNCIDSLVGL---QETIDGLQDD 376
QY 100 RII-----DPSLFSSGPTVDTAMRNACVNTSCGASVSSYANTAGNYYYRSPNSLT 151
DB 377 AIIIVHSDSGHGGEYERYPKEMERFQVCRTNQLSGSKBELVN-----V 422
QY 152 YNNLINTNCVLTGK 165
DB 423 YDNTLYTDFLTK 436

RESULT 4
ID 027167 PRELIMINARY: PRT; 2395 AA.
AC 027167;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE 51B TYPE SURFACE PROTEIN.
OS Parametium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Pentastula;
OC Parametium.
OX NCBI_TaxID=5888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STOCK 51;
RA MEDLINE=95098630; PubMed=7800503;
RA Scott J., Leeck C., Forney J.;
RT "Analysis of the mitochondrial B type surface protein gene in
RT Parametium tetraurelia.";
RT Nucleic Acids Res. 22:5079-5084(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-STOCK 51;
RA Scott J.M., Leeck C.L., Forney J.D.;
RA Genetics 133:189-198(1993).
DR EMBL; U07603; AAB81947.1; -.
DR EMBL; L04795; AAA16710.1; -.
DR InterPro: IPR002895; Parametium_SA.
DR InterPro: IPR003659; PSI.

```



```

DB 39 SSCGOTSNOCNASSLIGYNGGLISNTYN-----NNNOYDMNSQ--YGNOMGTTGSGTYCT- 90
QY 80 WSYHVAIVSYKNASGVTEKRIIDPSLFS-----SGPV-----TDTAMRACVN 123
DB 91 -----TSYSCRGSETCMNNRCQSSYSNNMYNSQYSNCSGTCRYSRDCSGQMCN 142
QY 124 TSCGSASVSSYANTAGNYRSPSNY-LYDNNL-INTNCV 162
DB 143 GVC-----VAOYGTSTYNNAMYSSSTGNCTYDNTGCLNORCV 179

RESULT 7
ID 002364 PRELIMINARY; PRT: 927 AA.
AC 002364; P90955; O62351;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE HYPOTHEICAL 102.3 KDA PROTEIN T01D3.6 IN CHROMOSOME V.
GN T01D3.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Steward C.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS: ISOFORM T01D3.6A (SHOWN HERE) AND
ISOFORM T01D3.6B; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; 281537; CAB04280.1; JOINED.
DR EMBL; 281110; CAB04280.1; JOINED.
DR EMBL; 281527; CAB03262.1; JOINED.
DR EMBL; 281527; CAB04279.1; JOINED.
DR EMBL; 281110; CAB03263.1; JOINED.
DR EMBL; 281527; CAB03263.1; JOINED.
DR HSSP; P02671; 1F2D.
DR WormPep; T01D3.6A; CE12964.
DR WormPep; T01D3.6B; CE18164.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR002919; TIL.
DR Pfam; PF00147; fibrinogen_C; 1.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM000179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00186; FBG; 1.
DR SMART; SM00216; VWD; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 1.
KW Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;
KW Hydroxylation; Hypothetical protein; Repeat.
FT VARSPLC 405 416 MISSING (IN ISOFORM T01D3.6B).
SQ SEQUENCE 927 AA; 103413 MW; 71AEF0A61FC2B266 CRC64;

```

Query Match 9.0%; Score 89.5; DB 5; Length 927;
 Best Local Similarity 20.1%; Pred. No. 2.3; Mismatches 68; Indels 149; Gaps 16;

QY 4 VIPDVATLSLNFQI--KNQSGCTSTASSPCITFRYPVDCYARAHKMOI-----LMN 55
 DB 130 VCPDYASGFFCCNETKCKNCSGKNA-----DCYVANHQLCLCKPGYTARR 176

```

QY 56 NGYDCEKQ-----EYVGNLKAS-TGTCCVAMS----- 81
DB 177 NGRDCMKKVOACMSGDPPHYVTDGLRFDYQGTCPVFSQPCITLDPAPLYMSVRAKNEL 236
QY 82 -----YHVA-----ILVSKN-----ASGVTEKRIIDPSLFS-----S 109
DB 237 PGKGYHISQVSEVEVDLHNLTHVDGRSKTALVNGV---QVLTPWFPPKNTWTVRYS 293
QY 110 GP-----VTDAMRACV-----NTSCGSA-----SVSYANTAGNV 141
DB 294 GSTFTIENDOGVYVFTTYSNLCVQPDIPFENGATTLGAGINDGKLDVYNNKGSV 353
QY 142 Y-----YRSPSN-----SYLDNNLI-----NTNCLVTKFSLLS 170
DB 354 LAIKSSROFENNHNHADFMKTEDTWTIDKFLILRPGQENCINGOTLDNNTCVSTISLAQ 413
QY 171 GCS 173
DB 414 SCA 416

```

```

RESULT 8
ID 0950P7 PRELIMINARY; PRT: 672 AA.
AC 0950P7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HYPOTHEICAL 71.4 KDA PROTEIN.
GN D1044.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans sequencing consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Pauley A.;
RL "The sequence of C. elegans cosmid D1044.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; U00065; AL27237.1; -.
KW Hypothetical protein.
SQ SEQUENCE 672 AA; 71395 MW; 1717219E5C8375A2 CRC64;

```

Query Match 9.0%; Score 89; DB 5; Length 672;
 Best Local Similarity 24.1%; Pred. No. 1.8; Mismatches 88; Indels 30; Gaps 9;

QY 20 NOSC-----GSTASSPCITFRYPVDCYARAHKM-ROILMNGYDCEKQF--VY 66
 DB 208 NOOCVAINGLNQLGTVSYNSOCITLASPGQNCQSSQCIDNSVCMNOMCTCANNRILYV 267
 QY 67 GNKASTGTCCVAMSYHV--ALVSYKNASGVTEKRIIDPSLFSGPVT---DTAMN 119
 DB 268 GCVPTTSSICQOTQTLVNNQCVLLSIVGFTCIANQOCVGMANGTCQTNGATAMYG 327
 QY 120 ACVNNISGASASVSYANTAGNYR-----SPNSGLVYNNNLINTN--CVLTKFSLSSC 172

CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Goebye J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S.S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbawani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale A., Baxendale B., Bhandari D., Bolshakov S.,
RA Beeson K.T., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Stimpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svistkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Glods R.A., Myers E.W., Rudin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RC Science 287:2185-2195(2000).
RL -1 SIMILARITY: CONTAINS 1 SET DOMAIN.
DR EMBL: AEO03704; AAF55041.2; -
DR HSSP: P20393; 16Y;
DR FlyBase; FBgn0003862; trx.
DR InterPro; IPR003889; FYRIC_C.
DR InterPro; IPR003888; FYRIC_N.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF006628; PHD; 2.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 4.
DR SMART; SM00508; PostSET; 1.
DR SMART; SM00184; RING; 3.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS50280; SET; 1.
SQ SEQUENCE 3726 AA; 400095 MW; E3DDB8F062BD7796 CRC64;

```

Db 178 SSGRSSGSSPDGNSGASSDGA-----SSGISCGKSTAKSTEASSGKLAKTTGA 225
Qy 74 GTCCVAAHYHAILVYKKNASGVTEKRLID-----PSLFSSCPVVDIARNACVMTSC 126
Db 226 GTCSGA-----KSSKASGGTSEATTSGLSGACKALFVAIPATSTGLACALVSPG- 276
Qy 127 GSASVSSYANTAGNVYRSPNSLYLNINLINTNCVLTKEFSLSGC-----SPSPAPDV 180
Db 277 GSSGCGFPISALLRRKKNKKNLNLARGEVMLPSTKQLKOLNSPVVDNPSPPPT 336
Qy 181 SI 182
Db 337 AS 338

RESULT 14
Q9GTE0 ID Q9GTE0 PRELIMINARY; PRT; 619 AA.
AC Q9GTE0;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE NUCLEAR RECEPTOR NHR-66 (FRAGMENT).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Bogan A., Maina C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
RT compatibility with the ligand-binding domain fold."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; AF273819; AAG15168.1; -.
DR HSSP; P20393; 1AGY.
DR InterPro: IPR000536; Hormone_rec_119.
DR InterPro: IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec_1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; znf_C4; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
KW NON_TER 1
SQ SEQUENCE 619 AA; 69082 MW; 6580E456F0E46D48 CRC64;

```

Query Match 8.48; Score 83.5; DB 5; Length 619;
 Best Local Similarity 21.58; Pred. No. 5.7;
 Matches 49; Conservative 20; Mismatches 66; Indels 93; Gaps 11;

```

Qy 23 CGTSTASSPCTTFRYPYDGCYA-RAHKMKROLIMNNGYDCEKQFYGNLK-ASTGTCVAV 80
Db 158 CGTSTG-----IHFGVDACACSAFRRRYVNLKNDYSCNKGKCTYVKDGSAGQKCRAC 212
Qy 81 SYHVAIVSY-----KNASGVTEKRIIDPSL-----FSSGP- 111
Db 213 RFRKCISSGMDKNSVQHRDAIGKYSAGV--KRELSPDAFEPEPSAKYSTVSEPTSSGPS 270
Qy 112 -----VDTARNACVNTSCSASVSS-----YANT 137
Db 271 GGFNONVSSPAGIPRVPSLTFTTQASTCMNSACGOKSVLHLLICRONFLTEOROLFYAGC 330
Qy 138 AGNYYRSPNSLYDNILINTNCVLTKEFSLSGCSPSPADVSSCGF 185
Db 331 LGD-WFKRPSS---IENOTL-----SELTDFSSCMF 357

```

RESULT 15

```

Q9GTD9 ID Q9GTD9 PRELIMINARY; PRT; 624 AA.
AC Q9GTD9;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE NUCLEAR RECEPTOR NHR-66 (FRAGMENT).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Bogan A., Maina C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
RT compatibility with the ligand-binding domain fold."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; AF273820; AAG15169.1; -.
DR HSSP; P20393; 1AGY.
DR InterPro: IPR000536; Hormone_rec_119.
DR InterPro: IPR001628; zf-C4.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; znf_C4; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
KW NON_TER 1
SQ SEQUENCE 624 AA; 69495 MW; DC5DIDEE608ADAF0 CRC64;

```

Query Match 8.48; Score 83.5; DB 5; Length 624;
 Best Local Similarity 21.58; Pred. No. 5.7;
 Matches 49; Conservative 20; Mismatches 66; Indels 93; Gaps 11;

```

Qy 23 CGTSTASSPCTTFRYPYDGCYA-RAHKMKROLIMNNGYDCEKQFYGNLK-ASTGTCVAV 80
Db 163 CGTSTG-----IHFGVDACACSAFRRRYVNLKNDYSCNKGKCTYVKDGSAGQKCRAC 217
Qy 81 SYHVAIVSY-----KNASGVTEKRIIDPSL-----FSSGP- 111
Db 218 RFRKCISSGMDKNSVQHRDAIGKYSAGV--KRELSPDAFEPEPSAKYSTVSEPTSSGPS 275
Qy 112 -----VDTARNACVNTSCSASVSS-----YANT 137
Db 276 GGFNONVSSPAGIPRVPSLTFTTQASTCMNSACGOKSVLHLLICRONFLTEOROLFYAGC 335
Qy 138 AGNYYRSPNSLYDNILINTNCVLTKEFSLSGCSPSPADVSSCGF 185
Db 336 LGD-WFKRPSS---IENOTL-----SELTDFSSCMF 362

```

Search completed: June 27, 2002, 21:58:26
 Job time: 659 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 20:16:15 ; Search time 3055.3 Seconds
(without alignments)
7397.198 Million cell updates/sec

Title: US-09-727-769A-7
Perfect score: 1080
Sequence: 1 agttaaataacacacacac.....atcagcaggtgctgtatat 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBml: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_on: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
------------	-------------	-------	--------	-------	-------------

1	1080	100.0	1080	6	AX113623	AX113623 Sequence
2	1080	100.0	1380	1	AB046594	AB046594 Chryseoba
3	555	51.4	555	6	AX113621	AX113621 Sequence
4	527.4	48.8	1080	6	ARI59962	ARI59962 Sequence
5	527.4	48.8	1080	6	E59342	E59342 Novel prote
6	344.2	31.9	555	6	ARI59957	ARI59957 Sequence
7	344.2	31.9	555	6	E59337	E59337 Novel prote
8	60	5.6	7218	6	166494	166494 Sequence 14
9	45.4	4.2	7319	6	AX346947	AX346947 Sequence
10	44.4	4.1	5993	8	AB018378	AB018378 Glycine m
11	44.4	4.1	13776	1	STATSC	L1198 Staphylococ
12	44.4	4.1	46445	1	AF051917	AF051917 Staphyloc
13	44.4	4.1	197405	8	ATCHRIV40	AL161540 Arabidops
14	44.4	4.1	202860	8	ATFCA2	297337 Arabidops
15	44.4	4.1	1182	9	BC016000	BC016000 Homo sapi
16	43.6	4.0	20992	3	LEIKEMAX	M10126 Leishmania
17	43.6	4.0	43049	9	HSDJ87H18	AL121846 Human DNA
18	43.6	4.0	117445	2	AL160263	AL160263 Homo sapi
19	43.4	4.0	2324	5	XELNRP1B	M34895 X.laavis ne
20	43.2	4.0	92633	2	PFMAL4P1_3	Continuation (4 of
21	43	4.0	145640	2	AC018361	AC018361 Homo sapi
22	43	4.0	158254	2	AC098659	AC098659 Rattus no
23	42.8	4.0	1141	6	AX083744	AX083744 Sequence
24	42.8	4.0	1141	6	AX083744	AX083744 Sequence
25	42.8	4.0	122051	8	AC005970	AC005970 Arabidops
26	42.6	3.9	88064	2	AC022707	AC022707 Homo sapi
27	42.4	3.9	17131	6	AX345955	AX345955 Sequence
28	42.4	3.9	219244	2	AL670771	AL670771 Mus muscu
29	42.2	3.9	6314	6	AX281272	AX281272 Sequence
30	42	3.9	125020	9	AF429315	AF429315 Homo sapi
31	41.8	3.9	1857	6	ARI57645	ARI57645 Sequence
32	41.8	3.9	77835	2	PFMAL13P2_3	Continuation (4 of
33	41.8	3.9	181596	2	AC068158	AC068158 Homo sapi
34	41.6	3.9	5987	6	AX345465	AX345465 Sequence
35	41.4	3.8	5884	6	AX347067	AX347067 Sequence
36	41.4	3.8	55984	3	U97196	U97196 Caenorhabd
37	41.4	3.8	56511	3	AC084467	AC084467 Caenorhab
38	41.2	3.8	62423	2	AP004554	AP004554 Oryza sat
39	41.2	3.8	82812	9	HSJ764022	AL096799 Human DNA
40	41.2	3.8	102237	2	AC099322	AC099322 Danio rer
41	41	3.8	7508	6	AX344203	AX344203 Sequence
42	41	3.8	73345	8	AJ294725	AJ294725 Astasia 1
43	41	3.8	84563	5	AC096885	AC096885 Dario rer
44	41	3.8	125020	9	AF429315	AF429315 Homo sapi
45	41	3.8	136198	2	AL392004	AL392004 Homo sapi

ALIGNMENTS

RESULT	1	LOCUS	AX113623	Sequence	7 from Patent EP110696.	DNA	Linear	PAT	30-MAY-2001
--------	---	-------	----------	----------	-------------------------	-----	--------	-----	-------------

DEFINITION	AX113623	ACCESSION	AX113623	VERSION	AX113623.1	GI:13939800	ORGANISM	Chryseobacterium sp. No. 9670.	Chryseobacterium sp. No. 9670.	Bacteria; CFB group; Flavobacteriia; Flavobacteriaceae;	Chryseobacterium.	REFERENCE	1 (bases 1 to 1080)	AUTHORS	Yamaguchi, S.	TITLE	Protein-deamidating enzyme, microorganism producing the same, gene encoding the same, production process therefor, and use thereof	JOURNAL	Patent: EP 110696-A 7 13-JUN-2001;	Amano Enzyme Inc. (JP)	FEATURES	location/Qualifiers	1..1080	/organism="Chryseobacterium sp. No. 9670"	/db_xref="taxon:161487"	61..1023	/note="unamed protein product"
------------	----------	-----------	----------	---------	------------	-------------	----------	--------------------------------	--------------------------------	---	-------------------	-----------	---------------------	---------	---------------	-------	--	---------	------------------------------------	------------------------	----------	---------------------	---------	---	-------------------------	----------	--------------------------------

```

/codon_start=1
/transl_table=11
/protein_id="CAC37857.1"
/db_xref="GI:13939801"
/translation="MKNLFLSMAAFVTVLTFNFNSCADSNGNOEINGKERLSYNDSKLKD
FGKTVPGVIDEENMIRKVSFMLTAQFYEIKRTKNEQYIGALRQAVNSESFVHFLK
NSNIGKVESASPEDVYFPTILTRKEKGGQNTKASVIPDAVLTNSLFENQIKNSCST
STASPCITERYYPDVCGRARAHKRRQILMNNNGYCEKQFVYGNLKSAGTCCVAMSVY
VAIVSYKNAAGVTEKRIIDPSLFSQSPMDTARMACVNTSCGSASVSYAANTAGNAVY
YRPSNSLYLDNNLINTNCVLTFRFLSLSGSPSPADVSSCGP"
466.1020
mat_peptide
/product="unnamed"
BASE COUNT      371 a      204 c      203 g      302 t
ORIGIN

```

Query Match	100.0%	Score 1080;	DB 6;	Length 1080;
Best Local Similarity	100.0%;	Pred. No. 7,2e-236;		
Best Local 1080; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	agtaaaataacccaacccaacttaaaaaaacccacttaaacatracaaatracaaat	60
Db	1	AGTTAAATTAACCAACCACTTAAACAAAACCTACCAATTAACATAAATTAACAATTAAT	60
QY	61	atgaaaaatctcttttltta tcaatgtagtccttttgacgcgtcttaacttcaatccggt	120
Db	61	ATGAATAATCTTTTTTTTCATGATGATGCGCTTGTGACGCGTCTTAATTAATTCGTGT	120
QY	121	gccgaatcccaacgcggaaatccaggaaatcaacggaaaaggaaaaactaagtgttaattctc	180
Db	121	GCCGATTCACACGGGAATCAGAAATCAACGAAAGGAAAAACAAGTAAATGATTCCT	180
QY	181	aagctgaaaagaatttcggaaaagacgttacccgttaaggatagacgaaaagaacggaatga	240
Db	181	AAGCTGAAGAATTCGGAAGAACATCTACCCGTAGGGATAGACGAAGAAACGGAAATGTA	240
QY	241	aagbgtlcaattatgttaactctgcgcaattctatgaaatlaagccgacccaagaataatgag	300
Db	241	AAGGTGATTTAATGTTAATCGCGCAATTCATGAATAATTAAGCCGCAACCAAGAAATGAG	300
QY	301	cagatatacggaaatgcttagacagcgtgttaagaaatgaatccctgtacaaatcttcta	360
Db	301	CAGTATATCGGAATGCTTAGACAGCGCTTTAAGAAATGAATCTCCGTACATTTCTTGA	360
QY	361	aagccttaacgaatgaataagaaaagtggagctctgcaagctccgaaagacgaaatgat	420
Db	361	AAGCTATATACCAATGAATAATGAAAAATGGAGTGTGCAATCCGGAAAGACTAAGATAT	420
QY	421	tttaaaacgatactctgacaaaagaagtaaaaggccaacaaataaatctgcagtgtaatt	480
Db	421	TTTAAACGATCTCGACAAAAGATAAAGGCAACCAATTAATTTGGCAGTGAATTT	480
QY	481	ccctgtagtgaatlaataataattctttattcaataataataagaatcagctctgcgtacc	540
Db	481	CCTGATGATGCTACATTTAAATTTCTTTATTCATCAATCAATAAAGAAATCACTCTGGCGTAC	540
QY	541	tcatacgcgtctccaccaatgcatacaatctcagataatccgttagaagcgtatctatgcaaga	600
Db	541	TCTAGCGGCTCTCACCATGCATCAATTCATGATATCTCTTAGACGATGTTATGCAAGA	600
QY	601	gcccaataagatgagacaataatcttaatgaaacaacgcgtatagaactgltgaaanaaactltgla	660
Db	601	GCCCATTAAGATGAGACAATACTTAATGAACAACGCGCTATGACTGTGAATAAACAAATTTGTA	660
QY	661	tacggaaaacctaaaggatctcaacacggaaactctgctgtgtggtgtgagctaccacgtttga	720
Db	661	TACGGAAAACCTPAAGGCTCAACAGGAACCTTGCTGTGTGGGTGAGAGCTACACAGTTGCA	720
QY	721	ataatgtgaagctataaaatctgctcccgagtaacgaaaaaagaatataatgtatcccttca	780
Db	721	ATATTGTGTAACCTATAAATAATGCTTCGCGAGTAAACGAAAAAAGAAATTAATTATGTCCTTCA	780
QY	781	ctattttcaagcggctcctgtatacagatacagcatggaagaaacgcttgcttcaacacctct	840

Db	781	CTATTTTCAAGGGGCTCTGTAAACAGATACAGATGAGAAAGCTTGGCTTAAACCTCT	840
Oy	841	tggagatctgcataccggtttccctcttaatgctaactgcaggaaatglttattacagaagt	900
Db	841	TGCGGATCTGCATCCGTTGCTCTTATCTATATCTGAGGAAAGTTTATTACGAAGT	900
Oy	901	ccatglaattcttactcgttatgacacaacatctgtaacaaataccaactgttactgaactaa	960
Db	901	CCTATTAATTTCTTACCTGTATGACAAACAATCGATCATACCAATGTGTACTGACTGAAA	960
Oy	961	ttttacagccttcgcgagtgtrctccctcaactgcagacggatgtaattcgaagctgtagatt	1020
Db	961	TTTTACATGCTTTTCCGGAGTTTCTCTTCCACTGCACCGAGATGTATCCAGCTGTGAGATTT	1020
Oy	1021	taattaattgtaaaltttcaagcacctgcataattacagaacbaagcagggtgcgttlatat	1080
Db	1021	TAAATTAATTTGATTAATTTTACAGACACTCTCTATTTTACGAATTCACACAGCTGCTGTAAAT	1080

RESULT	2	
LOCUS	AB046594	1380 bp DNA linear BCT 27-Apr-2001
DEFINITION	Chryseobacterium proteolyticum prga gene for protein-glutaminase, complete cds.	
ACCESSION	AB046594	
VERSION	AB046594.1	GI:12597204
KEYWORDS		
SOURCE		
ORGANISM	Chryseobacterium proteolyticum (strain:9670) DNA. Chryseobacterium proteolyticum Bacteria; CFb group; Flavobacteriia; Flavobacteriaceae; Chryseobacterium.	
REFERENCE	1 (sites)	
AUTHORS	Yamaguchi,S., Jeenes,D.J. and Archer,D.B.	
TITLE	Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme that deamidates glutaminyl residues in proteins purification, characterization and gene cloning	
JOURNAL	Eur. J. Biochem. 268 (5), 1410-1421 (2001)	
MEDLINE	21153247	
REFERENCE	2 (bases 1 to 1380)	
AUTHORS	Yamaguchi,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-JUL-2000) Shotaro Yamaguchi, Amano Pharmaceutical Co. Ltd., Research and Development; Suel, Kagamiyama, Gifu 509-0108, Japan (E-mail:LDV01447enifty.ne.jp, Tel:81-583-79-1220, Fax:81-583-79-1232)	
FEATURES	Location/Qualifiers	
source	1..1380	
	/organism="Chryseobacterium proteolyticum"	
	/strain="9670"	
	/db_xref="taxon:118127"	
sig_peptide	181..243	
	/gene="prga"	
	181..1143	
	/gene="prga"	
	/codon_start=1	
	/transl_table=11	
	/product="protein-glutaminase"	
	/protein_id="BAB21508.1"	
	/db_xref="GI:12597205"	
	/translation="MKNLFLSMAFVTLTFNSCADSNGNOINGEKLSVNSDKLKD FGKRVPGVIDEENMKIVSFMLTAQFELPTKENBOYIOMLRQAVNESPVHFLKRP NSNIGVIGASPEDVRYFKTILTKREYKGGONKLSAVIPVATLNSFLNFKNOSGT STASPCITFRYVDGCVARHAKMKROITLMNNGYDCEQFYVGNLKAATGTCVAMSH VAIVSYKNAKSGVTEKRIIDRHSFLSSCPATDTRARRNCVUTSCGASVSSTANTAGNV YRSPNSLYLYDNMLINTNCVLTFRFSLSSCPSPADVSSCGP"	
	181..1143	
	/gene="prga"	
	586..1140	
	/gene="prga"	
	/evidence=experimental	
	/product="unnamed"	
BASE COUNT	472 a 247 c 247 g 414 t	


```

Db      541  TCCAGCTGTGATTT 555
          |||
RESULT  4
LOCUS   AR159962
DEFINITION Sequence 11 from patent US 6251651.
ACCESSION AR159962
VERSION   AR159962.1 GI:16222841
KEYWORDS
SOURCE   Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Yamaguchi,S. and Matsura,A.
TITLE    Protein-deamidating enzyme, gene encoding the same, production
          process thereof, and use thereof
JOURNAL  Patent: US 6251651-A 11-26-JUN-2001;
FEATURES
          Location/Qualifiers
source   1..1080
BASE COUNT 354 a 210 c 205 g 311 t
ORIGIN

Query Match 48.8%; Score 527.4; DB 6; Length 1080;
Best Local Similarity 70.9%; Pred. No. 5,5e-110;
Matches 715; Conservative 0; Mismatches 291; Indels 3; Gaps 1;

Qy      17  caacttaacaaacacaccatlaactcaaatcaattatattgaataatctttt 76
Db      17  CAATTAAAGTTCACTAAACTAAACACCAAAATATAAACTAAGAAAAATTTCTGT 76

Qy      77  tatcaatgatggccttctgacgcgtcttaactttaactccctgtgcgaatccaacgga 136
Db      77  TATCCATGATGGCATTCTCGTACGATTTCTGTCAATCCCTCGTCAAGTTCAAGGCCA 136

Qy      137  atcaggaatacaacggaagaacaaactaagtgttaattcaattcaagctgaagattcg 196
Db      137  ACCAGACCCGCGATCTTGTCCGCTAAAGAAATCTAACGAACTCCCTA--TGAAAGATTTCG 193

Qy      197  gaaagagcttaccggttaaggatagacgaagaacgaagatgtaagtgtaattatgt 256
Db      197  GTAAGACGCTTCGGTAGGGATTGAAGAAAGATGAAATTTAAATCTCATTTATG 253

Qy      257  taactgcgcaattctatgaataatgaagccgaacaaagaataagcagatatcgsaagtc 316
Db      257  TTAAGTCCCAAGCGATATGAATTTGGGACAGTAAGAAATGACAGGTTATTTCCATGA 313

Qy      317  ttagaagaagctgttaagaatgaatctctgtacacatttcttaaaagcctaataagcaatg 376
Db      317  TCAGACAGCGCTGTGAGATGAATGAACCTCCGCTTCAATGTTTCTTAAGTCAACACCAATA 373

Qy      377  aaatgagaagaatgtagctgtcgaagtcgcgaagaagtaagatatttaaaagatcttga 436
Db      377  AATATGCAAAATAGAAAAAGCAACAGATGATGACATCCGTTATTTTAATGTGATTTGA 433

Qy      437  caaaagaagtaaaaggcaaaccaataaattgycgagtgtaattctctgtatgtagctacat 496
Db      437  ACAAGCAAGAGAGAGGTGAAGCAACAAAGCAGTCAAGTTATTTCTGATCTGGCAACGC 493

Qy      497  taattcttattcaatcaataaagaatcagctctgcgtaaccttaacgagcgtctcaac 556
Db      497  TGAACAGTTATTTTACCCGATCAAAACACAGCTTGGCGAATCTTCTACAGCATCTTCTC 553

Qy      557  catgatcatcatcagatatcctgtagagagtgtaattcaagagagcccaataagatgaagc 616
Db      557  CTTGATTCACCTTCAGATATCCGGTTGACGGATGTTATTCACAGGGCTCAAAAATGAGAC 613

Qy      617  aaatcttaatgaacaaagcgtatgactgtgaaaaaaatcttatacgcgaacaaactaaag 676
Db      617  CAATTAAAGTTCACTAAACTAAACCAAAATATAAACTATGAAGAAAAATTTCTGT 76

Qy      614  AATCTCTATGAACCGCGCTATGACTGTGAAGAACAGTTCGTATATGTATATCTGAGAG 673

```

```

Qy      677  catcaacagaactctgtctgtgagctggagctaccacgttgcgaatattgtaagctata 736
Db      674  CTTCTACAGGACATGCTGTGATCAWGGTATATATCAGTAGCAATTTGGTAACTTCA 733

Qy      737  aaaatgctccggagtaacggaaaaagaattatgtatcccttaactatttcaagcgctc 796
Db      734  AAAATGCTTCAGCAATTTGTTGAAGAAAGAAATGATATGATATTCATTATTCACGGGTC 793

Qy      797  cgttaacagatacagcatlgagaaacgcttgcgttaeacaccttgcgagatcgtacccg 856
Db      794  CTGTACACATTTCTGCATGAGAGCTGCATGACCAACACACAGCTCGGATCTGCTCTG 853

Qy      857  ttccctctatgcttaactatgcaagaaatggttatatacagaagcttagtaattctacc 916
Db      854  TATCTTCTACCGCAATFACGAGCAAGAAATGTTACTACAGAAAGTCGCGAGTTCAATAC 913

Qy      917  tgtatgaacaacatcgtatcaataaccactgtgtactgtacttaatttcaacgcttcgcg 976
Db      914  TGTATGATTAACAACATATGTAATFACCAATTTGTATTAACATATTTCTATCCCTTTCAG 973

Qy      977  gatgtctcctcaccctgcacgcgagatgtataccagctgtgatttaatt 1025
Db      974  GATGTTCTCTTCCCTCCACGACCAAGTGTAGCAAGCTGTGATTTTAAAT 1022

RESULT  5
LOCUS   E59342
DEFINITION Novel protein deamidation enzyme, gene encoding it, process for
          producing the same, and utilization thereof.
ACCESSION E59342
VERSION   E59342.1 GI:13023309
KEYWORDS JP 2000050887-A/8.
SOURCE   unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Shofaro,Y.A.M.M.
TITLE    Novel protein deamidation enzyme, gene encoding it, process for
          producing the same, and utilization thereof
JOURNAL  Patent: JP 2000050887-A 8-22-FEB-2000;
          AMANO PHARMACEUT CO LTD
          OS Chryseobacterium gleum
          PN JP 2000050887-A/8
          PD 22-FEB-2000
          PR 04-JUN-1999 JP 1999158703
          PI SHOTARO YAMAGUCHI,AKIRA MATSUURA
          PC C12N15/09,C12N9/80//A21D2/36,A23C9/154,A23C11/06,A23J3/00,PC
          A23J3/10,
          PC A23J3/16,A23J3/18,A23J3/34,A23L1/16,A23L1/23,A23L1/317,PC
          (C12N9/80,C12N1:01),(C12N9/80,C12N1:20),C12N15/00 CC
          FH Key Location/Qualifiers
          FT CDS (61)..(1020)
          FT mal-peptide (463)..(1017).
FEATURES
          source 1..1080
          /organism="unidentified"
          /db_xref="taxon:32644"
BASE COUNT 354 a 210 c 205 g 311 t
ORIGIN

Query Match 48.8%; Score 527.4; DB 6; Length 1080;
Best Local Similarity 70.9%; Pred. No. 5,5e-110;
Matches 715; Conservative 0; Mismatches 291; Indels 3; Gaps 1;

Qy      17  caacttaacaaacacaccatlaactcaaatcaattatattgaataatctttt 76
Db      17  CAATTAAAGTTCACTAAACTAAACCAAAATATAAACTATGAAGAAAAATTTCTGT 76

Qy      77  tatcaatgatggccttctgacgcgtcttaactttaactccctgtgcgaatccaacgga 136

```



```

Db      77  TATCCATGATGGCATTCGTGACGATTCCTGTCATTCATGCGCTGCTGAGATTCAGTGCACA 136
QY      137  atagaagaatcaacggaagaagaactaagttaattgcttaagctgaagaattcg 196
Db      137  ACCAGAGCCCGCAATCTTGTGCTAAGAAATCTTAACGAAAGCGGCA--TGAAGATTTTCG 193
QY      197  gaaagactgacgcgtgagatagacgaagaagaacgaatgataaagtgatcaattatgt 256
Db      194  GTAAGACTGTTCCGCTAGGAGATTCGAAAAAGATGGAATAATTAAATCTCATTTATGG 253
QY      257  taactgcgcaattctatgaattaaagccgaaccaagaanaatgagcaglatatcggaaatgc 316
Db      254  TTACTGCCCGCCGCTATGAAATTCGGGAGACAGTAAAGAAATGACAGTTATATTTCATGA 313
QY      317  ttgagacggtgttaagaatgaattctctgttaacatttcttaagccctaagcagatg 376
Db      314  TCAGACAGGCTGTGAAATGAATGAACTCCGCTCATGTTTCTTAAAGTCAACACCATATA 373
QY      377  aataagaaagtgagatctcgcaagtcgcgaagacgtaagataatttaaaacgaatccctga 436
Db      374  AAATTCGAAAAGTAAAGAAACCAACAGATGATGATCCGTTATTTAAATCTGTATCA 433
QY      437  caaagaaglaaaaggcaaaccaataaattgagcagtgtaattcctgagtgaatcat 496
Db      434  ACAAGCAGAGAGAGGTGAAAGCAACAAACAGACGATGTTATTTCTGATCTGCAACGC 493
QY      497  taattcttattcaatcaataaagaatcagctcttgagctcctcgaagcgctcccaac 556
Db      494  TGAACACTTTATTTATCCCAATCAAAAACCAAGCTGCGGAACCTTCAGACATCTTCTC 553
QY      557  cagcattcacatcagaatccctgtgaacgagatgtatgaagagcccaagaatgaac 616
Db      554  CTGTATACACTTCAGATATCCGCTTACCGATGTTATGCAAGGCTCACAATAATGAGAC 613
QY      617  aaatcttaatgaacaacgctatgactgtgaaacaattgtatcaggaacctaagaag 676
Db      614  AAATCCATTTGAACGCGCGCTATGACTGTGAAAAAGCAGTTGATGTTATGTTAATCTGAGAG 673
QY      677  caacaagaagaactgtgtgtgtgagcggtacacagcttgaatatttgtaagcata 736
Db      674  CTCTTACAGGAACATGTCGTGTATCATGGTATATACAGTACGATTTTGTGTAAGCTTCA 733
QY      737  aaaaatgcttcgagatgaacgaagaagaattatgaatccctcactatttcaacgagtc 796
Db      734  AAAATGCTTACGAGATGTGTGAAGAAAGATCATAGATCCCTCATTTATCTCCAGCGGCTG 793
QY      797  ctgttaacagatlaacagcagtgagaagaacgctgagcttaacacaccttgcgagatcgatcgcg 856
Db      794  CTGTAAACAGATTTGATGAGAGACTCATGTACCAACACAGCTGCGGATCTTCGCTCG 853
QY      857  ttctccttctgtcaataacgcggaagaatttatttacaagaagcccaagtaattctaac 916
Db      854  TATCTTCTTCCGCAATTAACAGCAAGAAATGTTTACTACAGAGAGCTGCAAGTTTCATTAC 913
QY      917  tgaatgaacaacaatctgataataacacacgctgactgaactaaatttcaactgcttcgcg 976
Db      914  TGTATGATTAACAACTATGTGAATATCAATGTGTATTAACATATTCTCATCCCTTTACAG 973
QY      977  gatgtctcttcaactgacgcggaatglatccagctgtgtaatttaatt 1025
Db      974  GATGTCTCTCTCCGCCAGACCAAGTGTAGCAAGCTGTGATTTTAAATT 1022

```

RESULT 6
 LOCUS ARI59957 555 bp DNA linear PAT 17-OCT-2001
 DEFINITION Sequence 5 from patent US 6251651.
 ACCESSION ARI59957
 VERSION ARI59957.1 GI:16222831
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 555)
 AUTHORS Yamaguchi, S. and Matsura, A.
 TITLE Protein-deamidating enzyme, gene encoding the same, production process thereof, and use thereof
 JOURNAL Patent: US 6251651-A 5 26-JUN-2001.
 FEATURES
 source Location/Qualifiers
 1..555
 BASE COUNT 159 a 122 c 112 g 162 t
 ORIGIN

Query Match 31.9%; Score 344.2; DB 6; Length 555;
 Best local Similarity 76.7%; Pred. No. 3, 2e-68;
 Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```

QY      472  agtgaattccctgtagtagcatataattcttattcaatcaataaagaatcagctc 531
Db      7  AGTGTATTCTCTGATCTGCAACCGCTGACAGATTTATTTACCCGATCAAAAACCAAGCT 66
QY      532  tgcgttacccctcagcggtctcaccatgcatcacatcagatatacctgtgaagagatg 591
Db      67  TCCGGAACCTTCTACAGCATCTTCTCTGTATCACTTCAGATATCCGTTGACGAGATGT 126
QY      592  tatgaagagcccaataagatgagacaactcttaatgaacaacgctatgaactgaaanaa 651
Db      127  TATGCAAGGGCTCACAAAATGAGACAAATCTTATTTGAACCGCGCTATGACTGGAAGAAG 186
QY      652  caattgtalacggaacccaataaaggcatcaacaggaactgtgtgtgagctgagctac 711
Db      187  CAGTTCGTATATGATATCTGAGAGCTTCTTAACAGAACATGCTGTATCATGGATATAT 246
QY      712  caagttgcaatattgtgaagctataaaaatgcttcgcgagtgaaaggaagaagaatatt 771
Db      247  CAGCTACCAATTTTGTGTAACCTTCAAAAATGCTTCAGGAATGTGTAAGAAAGATCATATA 306
QY      772  gatccctcacatttcaagcggtcgtgaacaagatacagatgagaaacgcttcgct 831
Db      307  GATCCTCATATTATTTCTCCAGCGGCTCGTGAACAGATTTCTGATGAGAGAGCTGATACAC 366
QY      832  aacacctcttcgagatcgtacatcgcttccctctatgaatcactcaggaagaatgttat 891
Db      367  AACCAAGAGCTCGGAGATTCGCTGTATCTTCCATCCCAATACAGCAGGAATGTTTAC 426
QY      892  taagaagctcctgtaattcttactgtatgacaacaatcgtacatatacacaatgctgta 951
Db      427  TACGAAGTCCGTCAGGTTTATTACTGTATGATTAACAACATATGTGAATTACCAATTGTGTA 486
QY      952  ctgactaatttcaacttccgcgttctcctcactcaactgacccgagatgataccaga 1011
Db      487  TTAACAATATTCTCATGCCCTTTCAGAGATGTTCTCTCCAGCAGCAACAGATGTAGCAAGC 546
QY      1012  tgtgagattt 1020
Db      547  TGTGGATTT 555

```

RESULT 7
 LOCUS E59337 555 bp DNA linear PAT 07-FEB-2001
 DEFINITION Novel protein deamidation enzyme, gene encoding it, process for producing the same, and utilization thereof.
 ACCESSION E59337
 VERSION E59337.1 GI:13023304
 KEYWORDS JP 2000050887-A/3.
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 555)
 AUTHORS Shotaro, Y. A. M. M.
 TITLE Novel protein deamidation enzyme, gene encoding it, process for producing the same, and utilization thereof

JOURNAL Patent: JP 2000050887-A 3 22-FEB-2000;
AMANO PHARMACEUT CO LTD
COMMENT OS Chryseobacterium gleum
PN JP 2000050887-A/3
PD 22-FEB-2000
PR 04-JUN-1999 JP 1999158703
PI SHOTARO YAMAGUCHI, AKIRA MATSURA
PC C12N15/09, C12N9/80//A21D2/36, A23C9/154, A23C11/06, A23J3/00, PC
A23J3/10,
PC A23J3/16, A23J3/18, A23J3/34, A23L1/176, A23L1/23, A23L1/317, PC
(C12N9/80, C12R1:01), (C12N9/80, C12R1:20), C12N15/00 CC
FH Key Location/Qualifiers
FT source 1.555
FT /organism='Chryseobacterium gleum',
FEATURES
source 1.555
Location/Qualifiers
/db_xref='taxon:32644'
BASE COUNT 159 a 122 c 112 g 162 t
ORIGIN

Query Match 31.9%; Score 344.2; DB 6; Length 555;
Best Local Similarity 76.7%; Pred. No. 3.2e-68;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 472 agtgaattcctgattgctacattcaattcttattcaatcaataagaatcagttct 531
DB 7 AGTGTATTCTCGATCTGGCAACGCTGAACAGTTTATTACCACGATCAAAAACGAGCT 66

QY 532 tgcggtaccctgaagcggtccctaccatgcatcattcagatcagatcctgtagaagatt 591
DB 67 TGGCGAATCTTACAGCATCTTCTCTCTGATACCTTACGATATCCGGTTGACGGATGT 126

QY 592 tatgcaagaagccataagaatgagacaaatcttaactgaacaaagcgctgactgtgtaaaa 651
DB 127 TATGCAAGGCGTCAACAAATGAGACAATTCGATTTGAACGCCGCTATGACGTGAAAG 186

QY 652 caattgtatacgaagaacctaaaggcatcaacagaacttgctgtgtggtgagctac 711
DB 187 CAGTTCGATATGTGAATCTGAGACCTTCTACAGGACATGCTGTATCATGGGTATAT 246

QY 712 cagcgttgaattgtagaagataaaaatgcttcgcgaagtaacgaaaaaagaattatt 771
DB 247 CACGTAGCAATTTTGGTAGGCTTCAAAAATGCTTCAGGAATTTGTGAAAAAGATCAAT 306

QY 772 gatccttcaactattcaagcggtctgttaacagatagcagatgagaacgcttgct 831
DB 307 GATCCTTCATATTCTCCAGCGGTCCTTAACAGATTCGATGAGAGCTGCATGTAC 366

QY 832 aacacctctgcgagatcgcgacgttcctcctctatgcttaactgcaagaaatgcttat 891
DB 367 AACACAACTCGCGATCGCTGTATCTTCCTACGCAATACAGAGGAATGTTTAC 426

QY 892 tacagaagctcagtaattcttactctgtagaacaacatcgtatcaatcaacaactgtta 951
DB 427 TACACAATTCGCTCAGTTCATTAATGATTAACAACATATGTAATATGTTGTGTA 486

QY 952 ctgactaaattcactcgtcttcgcgagatgtctcctcactacgctgacgagatgacagc 1011
DB 487 TTAACAATATTCCTCATCTTCAGAGATGTCTCTCCCGACGACCAAGTGTACAGC 546

QY 1012 tctggaattc 1020
DB 547 TGTGATT 555

RESULT 8
LOCUS 166494/c 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494

VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dornier, F., Schefflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source 1.7218
Location/Qualifiers
/organism='unknown'
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 5.6%; Score 60; DB 6; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.0014;
Matches 6; Conservative 231; Mismatches 141; Indels 0; Gaps 0;

QY 126 ttccaacgggaatcaggaatcaacggaagaaactaagtgtaattcgaattcgaattc 185
DB 1441 TTGCTACRR 1382

QY 186 gaaagattcggaaagactgacggttagggatagacgaagaacggaatgataaggt 245
DB 1381 RRR 1322

QY 246 gtcaattatgttaactgacgaattctatgaatgaaccgacaaagaagaagacgta 305
DB 1321 RRR 1262

QY 306 tatcggaatgcttagacagcgctgttaagaatgatactcctgtacacattcttcaagcc 365
DB 1261 RRR 1202

QY 366 taatgcaatgaataggaagatgagctgtcgaagtcgcgagacgtaagattctaa 425
DB 1201 RRR 1142

QY 426 aacgatlctgcgaagaagaagtaaaaggcaacataatgagcagtgtaattcctga 485
DB 1141 RRR 1082

QY 486 ttagctacattaaatc 503
DB 1081 RRRRRRRRRRRRRRRATC 1064

RESULT 9
LOCUS AX346947/c 7319 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2018 from Patent WO0200928.
ACCESSION AX346947
VERSION AX346947.1 GI:18494833
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
1 (sites)
REFERENCE Olek, A., Piepenbrock, C. and Berlin, K.
AUTHORS
TITLE
JOURNAL
Patent: WO 0200928-A 2018 03-JAN-2002;
Epidemiology AG (DE)
FEATURES
source 1.7319
Location/Qualifiers
/organism='synthetic construct'
/db_xref='taxon:32630'
/note='chemically treated genomic DNA (Homo sapiens)'
BASE COUNT 1946 a 263 c 1552 g 3558 t
ORIGIN

	exon	<2379..2463	/gene="GmENOD93"	
	intron	2464..3165	/gene="GmENOD93"	
	exon	3166..3227	/gene="GmENOD93"	
	intron	3228..3597	/gene="GmENOD93"	
	exon	3598..3673	/gene="GmENOD93"	
	intron	3694..3926	/gene="GmENOD93"	
	exon	3927..>4021	/gene="GmENOD93"	
BASE COUNT		2229 a 944 c 741 g 2079 t		
ORIGIN				

Query Match	4.1%	Score 44.4;	DB 8;	Length 5993;
Best Local Similarity	55.9%;	Pred. No. 5;		
Matches 105;	Conservative 0;	Mismatches 81;	Indels 2;	Gaps 1;

Oy	4	taaataaaccaaccacttaacaaaacctccatctaaactacaattacaatatcagc	63
Dd	1991	TAAAGATGACCAAGAAGTACTTAACAAGCCAGGTATTAAAGAAAATAAAAAAAGAAGATGA	2050
Oy	64	aaaaatccttttctaataatgatggcgcttlytgaccgctccaactttaattcctgcgc	123
Dd	2051	AACCTACCCTCCTTTTCATATATCATCTAGCGTGCTTATTAA--TTACAAACCTTAGCT	2108

Db	2109	TATTCGAAGGGATTTTGGAAACATGCATTAAGATCAACGTACCGCTAACGTAACTCA	2168
Oy	184	ctgaaga 191 	
Db	2169	GTGAACA 2176	
RESULT	11		
STATRSC			
LOCUS		13776 bp	DNA
DEFINITION			linear BCT 14-MAR-2000
ACCESSION			
VERSION		L11998	
KEYWORDS		L11998.1	GI:310606
		trsA gene; trsB gene; trsC gene; trsD gene; trsE gene; trsF gene;	
		trsG gene; trsH gene; trsI gene; trsJ gene; trsK gene; trsL gene;	
		trsm gene; trsn gene.	
		Staphylococcus aureus.	
SOURCE			
ORGANISM		Staphylococcus aureus	
		Bacteria; Firmicutes; Bacillus/Clostridium group;	
		Bacillus/Staphylococcus group; Staphylococcus.	
		1 (bases 1 to 13776)	
REFERENCE			
AUTHORS		Morton, T.M., Eaton, D.M., Johnston, J.L. and Archer, G.L.	
TITLE		DNA sequence and units of transcription of the conjugative transfer	
JOURNAL		gene complex (trs) of Staphylococcus aureus plasmid pG01	
MEDLINE		J. Bacteriol. 175 (14), 4436-4447 (1993)	
FEATURES		93322322	
SOURCE			
		Location/Qualifiers	
		1..13776	
		/organism="Staphylococcus aureus"	
		/isolate="pG01"	
		/db_xref="taxon:1280"	
gene		complement(161..347)	
		/gene="trsn"	
		complement(161..349)	
		/gene="trsn"	
		/note="putative"	
		/codon_start=1	
		/transl_table=1	
		/protein_id="AAV1951.2"	
		/db_xref="GI:7239722"	
CDS			

gene
CDS
/translation="MNNNEENSVEFFGKKKKVSLHLVDPDMKDEIITKYAQEKEDPNVS
QAGREILKKGLEQIKSNK"
/gene="trsa"
521..1495
/note="trsa"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71952.1"
/db_xref="GI:310608"
/translation="MDNITFRKRIEYFKRSIEQKKYSKLNTOGRFVCKDEYSTN
IKGIATDKENKLVAAEDREDKQKVLANKDNMYTLNLSGTYKEAQNFTGNTNY
EMVDELFIYNSKQALNTYDQDEYHNROKQLOYNLEQERKQRLDITPKNSNAK
EDLALSLGTATLAKGININISNVEKVTNEMSMVITENINGKSEHLISDFDSEIEKA
KELTVSPSMRLGFEFEEKNASGOIAMGTFEFGAGYDESTLENDPKVATGEGYSVS
FDKQNKDILTADTFEEAVLKSLEMKNDLSRSEIENQLEKKI"
1512..1829
/gene="trsb"
1512..1829
/gene="trsb"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71953.1"
/db_xref="GI:310609"
/translation="MIKKFSLTIVYAVFLSIVLSNITLGAENPGFKIEQGLQOYQFL
TGLIYAVGICAGVWIVLKKPLGIDPMVKNEFMRGVAVGAVGALVWLVPVYN
LFO"
1889..2296
/gene="trsc"
1889..2296
/gene="trsc"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71954.1"
/db_xref="GI:310610"
/translation="MSVDEYKDIMLDMIGEEKYKENIIPRNDASYKIFLNLNGKE
LAIVFPLLELTIAFGVTLFGMLNLITGFNVFVGLIFGLTIYGLLTIRPISKEN
IRMDITIKOSORFSRQKVYFYKSESGDDDV"
2283..2966
/gene="trsd"
2283..2966
/gene="trsd"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71955.1"
/db_xref="GI:310611"
/translation="MMFMRKKKNKKNRNEQOELINRYIKGGRNOSIOEDSLIAGFY
KDYVTKDGYISGGLNITGVTELTTFEONDLEEYSSFINLIDPMIKTITEPV
NIQNYIKHLVOVLDMKNNNNKPTRGOLLAGRLFLQERLAKGEMSTKKHYIPESE
KITENTVQDLDAVRLHEKINLNKEALQAFSHIGLETRPIYQREYLNQFALYDYK
GYNDYHLSN"
2981..4999
/gene="trse"
2981..4999
/gene="trse"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71956.1"
/db_xref="GI:310612"
/translation="MIMAFIKKKOEOVTNKHNSFQKLTVEVLDTSDVSITYPSWI
EKXSHIETGENTIKNLVVDYPOSYKAGYLSLKKNGNIQITKRIRANIERMADHL
NNSIKNTAEQRTDPRKNATIKREISSKKQDOKFLDEKTGEMWYMTYTLNDGY
EKIQALEKDVARTLRLLKTHPTNARESHVLPNRFLSAFTQOMDNTAGH
FEMFDESEIIDLPNTSVFGINKNTSDSLAADFNNKEKTLNMTIIGTSGGKSTLN
MRMLDNKRSIROFIIDPEDEFYITKYGGTYVNISSTSNIKINPEIRFSEVEEK
EDENNETTSOVLTFENNEHESQIDLVRSKIGKILKPPRVLKDEISOMEISLSTIR
QLYQDKGFKNAKLSDFKSEDPITLTLYNKLKIDLPKREYVLKDLTLILIEDYTHRG
TTTIFDGYTNIKLQNEIVTNLKLQTERKDVOSAAYLNI FSLWDEITKRRTETVLM

gene
CDS
TDEHFLATNEXSLDFYQAYKRIKYGGAIASTQOIKILIRTSQETIGSAIENSHT
KEFFGMDNVGDVVDVKGGLKRFSDQELSHLRKKKGGAALLLYGTORAFIRIDLQREET
RLMNKELEYETIYEEDVADVEPNVBEQLGLTDLIDALEBEELQAEMIE"
5011..6291
/gene="trsf"
5011..6291
/gene="trsf"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71957.1"
/db_xref="GI:310613"
/translation="WKIRNEGTLVGHNHLEYIIVNDLCHISDIDEIRDLAPNE
LEFEVKKDQKQEDGYPHIYNTDNEYSFLSAKESKALKLSIMDYLVLPDKDITV
MHPALFERNIEDIKIGFGHEIYLPKPINLEQYKLLIMSTLSQYTFDKYRNKFEV
LSKREDDHFNNAENETPEKLEIYRHELNQODYLLNEEKREKREKRYIRNLIIG
FVGTAIVLSISFMINKGOEOLSKINAOQROKRSYVENLVENLVQNDQAVGKKRD
DSFNKDIKTLKREKKEBELIELSSKTPYIYBDLYKEGKONKIRELAFNENNDTL
SLEKTIIDKDGTAESVSGTYEESKRFLALASAGEYVESAKDINKKLQNEVEEI
NKAELIQLKEEKDSTKDKKKEIODELEEK"
6309..7385
/gene="trsg"
6309..7385
/gene="trsg"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71958.1"
/db_xref="GI:310614"
/translation="MMSAIAGCITMLMKKALKSKITLKPIAVIGFLLIFINAYLV
GSTFKEINSEGEQKKEASTGTACTGSVDGSIPEKNAKGALBEGKDKVYI
AKKNIPKPLFAIVASESEWGRGANAKOKNPISIMGAPLOVPISIEGIDKGN
LYDLIISGLTTPPEKIGPKYAPVAGASNDPDLNSMWPTVYKIMKSPGKEAKCTES
SGGSGDGKGFDEKGFEPKPKSKYNGOSYPMGOCTWYHORREIGKRPVPLTMGNGD
MGDNAAQAGWEVSGKPKRAGRSVYKPMFGAPPYGHIMFEKVKKDGGIYSEANVK
GLGVSSRPSFAETQRMQFIYDK"
7395..7880
/gene="trsh"
7395..7880
/gene="trsh"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71959.1"
/db_xref="GI:310615"
/translation="MKRLTYIILFTLVLAGGQNSKSEKSHNEKTVDTTEBRND
MIKGLIEGYENDDYKGYKHKMDILSEKLOQAELOQNSAQTKGYKKEKDKILYQDY
DNDDELIYSIKLSITNKKKNSDYSMRKGIETLPEBKOKIDNVEELTORLVSMNQ
E"
7896..9998
/gene="trsi"
7896..9998
/gene="trsi"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA71960.1"
/db_xref="GI:310616"
/translation="MNTLICEKPSQAMDLSTVPAKKKKONGYMEISDBOLNVSGFLT
MAVGHVLAKPEQEDERYKKNSTPILLEDNDPOFKSDKKDOPNNIKIKREKNI
DEVIITADIPAREGINAIYKILNQLVYTKVITKRAMLTSKESSIRKAFKNILPREKT
YGFYKEGARELSDWLVGINLSRHPTKISREBNGVLIHIGVSSPTLMMYRNENNI
KGFGKGFYKVSATINKDEQEVKFLKPFSEBNDLHFLFENDITDLPLOGLVNDIE
KEIGYTMPKFEYDLSALOEDMNDKYSIAKRTLEIAOTLYEKKILTYPRDSRYTDED
KEMLLENIDVILKEITKINLNNELTNNSLINPSKIEDHYAILITGDNFNKVDLKEEI
NVKXSILQNVANRPDKQYETTTIEIAVKKLMPEVKKIITODNFKALLNKQKTSSE
TIPNEKNEVEDIEIDLLEKETTPPKRYTEKTLKAMNPLETELDEBGLSKLKVKG
LGPATRADIDLENLKNRYIOVQKNKIYITKNGIACLLEBHLBSKDLGQWERYL
NGISGECODDSFINITNEMIKKTIENEYKNESIOKVAKEVSTNNIAKCPACDNGY
LIDRSGFGCEFYNNNGCEFTIPKRLLEKSIPTPYVKALCESVTTTKLGFPSKSKGS
FDCKILKKNKLOPSFD"
10014..10478
/gene="trsj"

CDS

10014..10478
/gene="trsr"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA071961.1"
/db_xref="GI:310617"
/translation="MKGIFFLTGIMLVLIYLAQILAYALNTLSMTSNOKIFSDIS
OLFIHPAHPISTIIIGFVSDKNVTFEYGVATATITCIYILIKQLFEKSDGKEMQDYK
IAKHSKSHAPKFAIEGSELSFSGHYKKMKENDITNVVNSLDSLIEERSDE"
10475..12115
/gene="trsk"
10475..12115
/gene="trsk"
/note="putative"
/codon_start=1
/transl_table=11
/protein_id="AA071962.1"
/db_xref="GI:310618"
/translation="MTTLGELSKATSIINKNLVIODEFTKCPENRIIFYVGGSGYK
SAGVYIPWVYIINOOSIVYDPGSGVEYKTSNIRKMGQEDVRVNFENFLASDQNEF
DYIDKSDSYIAELIISAGDSKINKDWYTKASVGLNSLLYAKIEFEPEKTKT
LIKFIQNNKPNODEGSEYELDKRFNELSKDHPAESEYEGFAVEGRTKRSILITPAA

gene

CDS

Query Match	4.1%;	Score 44.4;	DB 1;	Length 13776;
Best Local Similarity	46.7%;	Pred. NO. 4.7;		
Matches 141;	Conservative	0;	Mismatches 161;	Indels 0;
				Gaps 0;

Oy	224	aagaaacaggaaatgataaagctgctcatctatgtaactcgcgaattcatgtaattaac	283
Db	8935	AAGAAATACCAAGATTAATTTTAATATGATTAACCTTAATACAGCTTGATTATTCCTT	8994
Oy	284	cgcaccaagaanaatlgagcaglatatcggaaatgcttagaagcctglttaagaaatgaatc	343
Db	8995	CTAAATATGAGAGATTCATTACGCATATTTATATACTGGAATATTTTAAACAAGTAGATT	9054
Oy	344	ctgtacacatttctctaaagccctatagcagaatgaatagaagtgagcgtctgaagtc	403
Db	9055	TAAAGAGAGAAATCAATGTTTATATTAAGATTAATACAAATGTGGCCATGAATTTTA	9114
Oy	404	cggagaagcgtlaagatattttaaagacatccctgacaaagaagltaaaggycnaaccaata	463
Db	9115	TGGATTAAGACATATATGAACACATCACTATATAAAATCCCTTAAAAAGCTTATGTTTG	9174
Oy	464	aattgycgagctgtaattccctgactgtatagctacatlaattcttataccaataaaga	523
Db	9175	AAGGAAGAAGTAAATATTATTCAGATATACAGATTTAAACCTTTGTTAAATTAACAAAAA	9234
Oy	524	at 525	
Db	9235	CT 9236	

RESULT	12				
AF051917		AF051917	46445 bp	DNA	circular BCT 30-SEP-1998
LOCUS		Staphylococcus aureus	plasmid pSK41,	complete	sequence.
DEFINITION		AF051917	LI9570		
ACCESSION		AF051917.1	GI:3676412		
VERSION					
KEYWORDS					
SOURCE		Staphylococcus aureus.			
ORGANISM		Staphylococcus aureus			
		Bacteria; Firmicutes; Bacillus/Clostridium group;			
		Bacillus/Staphylococcus group; Staphylococcus.			
		1 (bases 23598 to 41563)			
		Firth,N., Ridgway,K.P., Byrne,M.E., Fink,P.D., Johnson,L.,			
		Pallesen,I.T. and Skurray,R.A.			
		Analysis of a transfer region from the staphylococcal conjugative			
		plasmid pSK41			
		Gene 136 (1-2), 13-25 (1993)			
JOURNAL		9412390			
MEDLINE		2 (bases 1 to 46445)			
REFERENCE		Berg,T., Firth,N., Apisiridej,S., Hettiaratchi,A., Iealaporn,A. and			
AUTHORS					

TITLE	REFERENCE	AUTHORS	JOURNAL	MEDLINE	COMMENT	FEATURES
Skurray, R.A.	Complete nucleotide sequence of psk41: evolution of staphylococcal conjugative multiresistance plasmids	J. Bacteriol.	180 (17), 4350-4359 (1998)	98389645		
3 (bases 23598 to 41563)	Berg, T., Firth, N., Apisiridej, S., Hettiaratchi, A. and Skurray, R.A.	Direct Submission	Submitted (03-OCT-1993) School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia			
Sequence update by submitter	4 (bases 1 to 46445)	Berg, T., Firth, N., Apisiridej, S., Hettiaratchi, A. and Skurray, R.A.	Direct Submission	Submitted (02-MAR-1998) School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia		
On Oct 1, 1998 this sequence version replaced gi.405558.	Location/Qualifiers					
1..46445	/organism="Staphylococcus aureus"					
/db_xref="taxon:1280"	/plasmid="psk41"					
72..686						

sig_peptide
mat_peptide
CDS
gene
CDS
CDS

755. .2026
/transl_table=11
/product="unknown"
/protein_id="AAC61931.1"
/db_xref="GI:3676413"
/translation="MGNIIDFDHHEGYRKLPLPKDFKHYFKHSDKTRNHPIFGAMFNKKT
DYESNFENSKMNIIPKNEFYIHIOQEPENSPSPKXVFNKIDSAETIOFLRQVGSSENY
LENIIRDVGLIEVYALHSFKKGAMALYSOLELSDKELPIPLYLTOTNLKDELYTOD
LGEFDPIKGNHGEPLLVYKPRNLETNKKPKIKSLFRKVFNNI"
755. .2026
/note="Orf423"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC61932.1"
/db_xref="GI:3676414"
/translation="MDYNOKRRTYIIIEGIVLATILVIGIAYVILISLNNAKGVKEFE
KLVQNDNSTIAKLKLSNKEETLSENDANFVLFPKSKENDEQVNEQIOLKISNIDKAK
DNIVYGSITPDYOGNNLLITIKNGKTEFFIDLEFKPRLSVYKKEFNNGIYKVOOE
NAKAKVIASNDOTTEGLKFEFGIKSTEDTIEDSIHNGTILGKIDIDLSRGNNAKLYE
ADEDNOSWFYKVLKNNDSLIEKRSYKLLIDQDVYDSKRYGKYKVPTELSYVAICK
LEGEKFKYRINIRPNYDNKPONLKLSEKSOIDPDSKSKYKEXKAKSYIKETKTEL
NKAYKKRYKITSIDYVKNNSOLEKOLKQVYSKSAVAFKDYKINNVRDKNKIYIELS
KNNANVFQTKITLDYDVAISNFIEDSHN"
755. .862
/evidence="not_experimental"
863. .2023
/product="unknown"
complement(2068. .2484)
/note="Orf149"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC61933.1"
/db_xref="GI:3676415"
/translation="MKIINPMNPEPKYETDNRNIPREYLNPHIPKRGIVYKNAKFT
IPQYELIEQIHEDQKINMPLSDQOLQETNEVYSKISKNILTDLYKKGDIYIISI
OQYINKIDFNGFMILITNSNSNNKLIPLSISVSO"
2779. .3336
/gene="res"
2779. .3336
/gene="res"
/codon_start=1
/transl_table=11
/product="putative resolvase Res"
/protein_id="AAC61934.1"
/db_xref="GI:3676416"
/translation="MANIGYARVSTQSDLIQDIALSNFGCEKIEFSEKYSGRVNT
ELDKIDILREDDTLVYKIDRLRGTTTQOLIDLQOWLENNNIETQIIDMNIINRKDPNG

```
CDS
KMF7MSAPAELENLISERTKGLSESARARANGNGRKAMPQEKKEYIKHLVDTKEY
TGOETAKRTNVSRTVYRIALREFKK"
3618..4397
/note="Orf259"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC61935.1"
/db_xref="GI:3676417"
/translation="MEVVRNLONLEFKLNGEISSNKNKSGRPYYSPFKRNNRVCIPE
RQAPAIQKQYDMNDYIKAKASLSKIPLVKSTQYFHKELNIDSDNOQTAKIN
ELISGKSNKTKLDOSLPNEKLNLDLDTETLTERKSLDYPAPAKINMSDIMPFELE
KNKHFTLSALTIKNEPKHYKDFLNDYIENEKNDIDLTL"
4531..6147
/note="Orf538"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC61936.1"
/db_xref="GI:3676418"
/translation="MRTPPGILNNTDYELNSEFYVPASDKOMNDORDMTKSIVGN
LYVRNPFVNEYTOPYVDLDNSEKQYOTPHIISIDDCEDYNTLTLDYFPKRTKNG
YEKLVAIKSEBITVEDSDINEVDFTQODESNDPWTNTPEVDLVLSKGYSTHOPN
RESVEDSHITTENDITYNVSINPEKHLYLNDPEALPTEENDIRLNDNLTOEMK
TYDININFSYNDEFKLEDDSYIIDLVDYKONEGKFVKNLIEDIPNEIILKRLKON
INLPDERANVALAVYRRDLIEDFSDRNMLSDKVKDLKTEPDEINIKGLINNDIE
KALNNOALFNLNPNFRDNVLEVEYEKEIYDEKIPLSIEDREKVTYLDAR
NNFSNPEYETTKSSILAEENSLTVLDKAKYDLPEFAPVSNHLYKREMPYEGMKATL
PLHNTKROYKEREDEITLIGHFKNYITMELHGTGSSSEPKPFVNSDLNENSTAFE
DIINGSETHSINIISIKNELKQERNLDM"
6350..8074
/note="Orf575: similar to Lactococcus lactis prs01 lttrc"
/codon_start=1
/transl_table=11
/product="lttrc-like protein"
/protein_id="AAC61937.1"
/db_xref="GI:3676419"
/translation="MAVSGNKKKKKIDSIIEELNNKLDPERKNDKTYKEFLDTTAKF
HNSYINNIILLAGONPEATAVAGYKAMKNKEDROQAKASNLITAPLYOKRNVYIKD
DNGNIVRDNGNGPKTERKPVITGYRAHVFDVADKRGKPLTAKILAPVOKRNVYIKD
LIYNEFKYINIEETLPSVEEKEFLDPLTNSGAKYVSPKSEIYISDLSYDMFKT
LHHEVASHQLNNDIGKTOISEHRSKLKEIEESAIYVANTYGLDTSYSLIGYSGN
GHNISDDELKAHINKHISFAKTYTIEINSLEPFSYIDKLESENNKQVYDLSWTID
INLKNGDFKITIKGNLVNDYGLNEISSENSYENDPKYNIYKGTNNNSODQAIEI
IKHDDSLNDRYNEFQTYNRVYINNTTIFVODDNEENKRYKIHEDPIDGNIIEEKI
KLSPODEIYSENFENKSVYENGKLDTLROFVNSGVGLYELVNDVNIINTNETYSMN
KNOONGORSNMTAKLEHDTNENVYIDFKIKNSAGIKNISTYESNEEFKDKQNKIEIN
KQEBEIDI"
complement(8115..10112)
/gene="nes"
complement(8115..10112)
/gene="nes"
/codon_start=1
/transl_table=11
/product="Orf1 nickase Nes"
/protein_id="AAC61938.1"
/db_xref="GI:3676420"
/translation="MAAYHFOKPFVSKRANGOSATKSAVNSARIKDKPERNEFKDYN
KOCYSEILLPNMADDKRDKREYLMNKYKHDVENKRSOVAEITIGLENDPKNITE
LAKEFAESLSNEGMLVDLNIHKINEENHAHLICLRKLDKNEPEPRKRGDNYTRDW
NTRKEHNEMWRKRWENVQNRKLEKNGFSVSVASDKNQNIDLEPRKKGWARKAFED
TGKSSISKHNESIKRRNQOKIDKMFNEVDPMKSHKLNAFSYMNSDSTTLKNAKDL
KIYVPIPNMYKENERLYDKOTSLITDROBLINKIEDIEDROKKLESINEVEKQAG
IEFDKNYPDOSLNSDDEKIFITRTILNDROYLPANNLEIDVYKRIKPAOISLNTV
LGNRDIJSESTEKESNEFADKLSNLEKNNISFPDVLNKHGEGMDSLKIDYTYTKLE
VPRNAENTLEDTYDQIKELFTDDEDTYAFNEVDIKKQDLIDKTYHGTENTIEML
ETGNFIPRYSDERKYITEQVKLLQEKFFPKKNQKNDHFVEGALQKLLSEDPYSD
NNDLKNLQOESNEVDEISKDNIEEFYSGNLDLVDKETEYNNRQOAYGLINAGLDS
MIFNENIEFERREMPYINHOYKGNHKSQRHELEKKNRGMHNY"
complement(10221..10222)
/note="putative origin of transfer ntc site"
10316..10586
misc_feature
CDS
```

```
/note="Orf90"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC61939.1"
/db_xref="GI:3676421"
/translation="MADIQKMLEQEKIQEKIKKKKEYEFGMFPNKKRVNSAD
AKIKNKMLNLEDETRKEHNLERNNEIIEETKNSNEENKEINEININI"
10692..10925
/note="Orf77"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAC61940.1"
/db_xref="GI:3676422"
/translation="MILSTAGENAVHINPIFRPFLYFIMCVPMVALCTIFIKRISI
FLPKVCLYIGIALGITIVTIFLVLTITGIEMT"
11188..12228
/note="Orf346"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC61941.1"
/db_xref="GI:3676423"
/translation="MSNVYVALDQFNGCFVKKINDEXEVIJPSRIGKRTNENNLKGF
VDNKLDVSEFTINGNNDVLLFGNDLDTKTYTKDPTASTNRYDIKSKFDLVEGSIIGL
LAREVEEVVNVVITGMPNSIEIGTKQAKFEKLLNKSRLIEIDIAITINVKGVKIV
AOPMTGLDLNMEKNGKVPKAPTEGYSVLDSGSGTITIDTYONKRRVEESFVINKGT
IDFKYRLSHVSKSSEKASITPRMIEKLEKLYKCKCKLNQKTYIDPDEPKKODSLIE
VMSNETTVGINSIDRLITVVGGANIHFDLSHYSYDVEKADDSQPSNVKGYEKLQ
ELKKNVQESK"
12231..12560
/note="Orf109"
/codon_start=1
/transl_table=11
/product="unknown"
Query Match 4.1%; Score 44.4; DB 1; Length 4645;
Best Local Similarity 46.7%; Pred. No. 4.2;
Matches 141; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 224 aagaaacggaatgataaagcgtcattatcttaactcgcgaatctctgaataaagc 283
|||||
DB 32538 AAGAAATTAACTAAGATTATTTGAATTAAGATTAACTTAACAGCTTGATTAACTCTT 32597
|||||
QY 284 cgaccaagaataagcagcgtatcgcgaatgcttagacagcgtcttaagaatgatactc 343
|||||
DB 32598 CTTAAATTAAGATCTTTACGCAATTTTAATTAACCTGGAATGATTTTAAACAAAGTAGATT 32657
|||||
QY 344 ctgtacacattctctaagcctaataagcaatgaaataggaagctgtagcgaagtc 403
|||||
DB 32658 TAAAGAAAGAAAGAAATGCAATGTTTATTAAGATATATTACAAATGCGCCATGAAATTTTAA 32717
|||||
QY 404 cggagaagcgttaagatattttaaagcgtcgcacaaagaatgaaagcgcaaaccaata 463
|||||
DB 32718 TGGATTAAGAAAGAAATTAAGAAAGCACTACTATTAAGAAATCGCTGTAAAGCTTAATGTTTGG 32777
|||||
QY 464 aattgcgcgctgtaattcctgtagtgcacattaaatccttatccaatcaataaaga 523
|||||
DB 32778 AAGTGAAGGTAAATTAATTTCAAGATACGGATTTTAAAGCTTTGTTAAATAACAAAAA 32837
|||||
QY 524 at 525
DB 32838 CT 32839
RESULT 13
ATCHRIV40
LOCUS ATCHRIV40 197405 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 40.
ACCESSION AL161540
VERSION AL161540.2 GI:7268224
KEYWORDS
```

SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids.IT, Brassicales: Brassicaceae; Arabidopsis. 1 (bases 1 to 197405)
AUTHORS	EU Arabidopsis sequencing project.
JOURNAL	Direct Submission
COMMENT	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mps.biochem.mpg.de, mayet@mps.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV41 at the 5' end and an overlap with ATCHRIV41 at the 3' end. Location/Qualifiers 1. 197405
FEATURES	/organism="Arabidopsis thaliana"
source	/variety="Columbia"
	/db_xref="taxon:3702"
	/chromosome="4"
exon	/complement(8150..8280)
	/gene="At4g14790"
	/number=1
gene	8150..10349
	/gene="At4g14790"
gene	complement(join(8150..8280,8363..8531,8621..8698,8884..9006,9126..9246,9299..9539,9618..9739,9883..10162,10286..10349))
	/gene="At4g14790"
	/note="similarity to probable RNA helicase SUV3, ATP-dependent, yeast (Saccharomyces cerevisiae), contains ATP/GTP-binding site motif A (P-loop) AAI36-143; EF-hand calcium-binding domain AA383-395"
	/codon_start=1
	/product="RNA helicase like protein"
	/protein_id="CA878521.1"
	/db_xref="GI:7268225"
	/translation="MAISVYLRKVSALGISRVQLSESLNYELCSFYICPCVVSRRRLDFNVSVDGFLPNVSADKSLMRFHEPEFGDLLGLVTRNYSKSPKDFGTGTTSKDFDLCPTWYPIARKKRRVILHVGPTSGKYSALKHLEQSSGVCGLRLAEVAKRLNKANVPCDLITGOEKDVEGTHKAVEMADVTSYDCAIIDEI OASLARMLKKSRTPTCLGFAPFALLGIADELIHLCGPAVVPIVIEDILKVGDDVE VHYTERSLVPLKVPYSSSITKTCGLVTFPSKDIYAKKTIERAGKHLCSYVYS LPETRTAQTRENDETNDFVLVASDAIGMLNLNLSRIIFSTLQKYDSESTDLTV SEIKQIAGRAGRSQSKPPIGEVTCLEKEDLPLHSSLSKSPSPLLEANLMI"
intron	complement(8281..8362)
	/gene="At4g14790"
	/number=1
exon	complement(8363..8531)
	/gene="At4g14790"
	/number=2
intron	complement(8532..8620)
	/gene="At4g14790"
	/number=2
exon	complement(8621..8698)
	/gene="At4g14790"
	/number=3
intron	complement(8699..8883)
	/gene="At4g14790"
	/number=3
exon	complement(8884..9006)
	/gene="At4g14790"
	/number=4
intron	complement(9007..9125)

	/gene="At4g14790"
	/number=4
exon	complement(9126..9246)
	/gene="At4g14790"
	/number=5
intron	complement(9247..9298)
	/gene="At4g14790"
	/number=5
exon	complement(9299..9539)
	/gene="At4g14790"
	/number=6
intron	complement(9540..9617)
	/gene="At4g14790"
	/number=6
exon	complement(9618..9739)
	/gene="At4g14790"
	/number=7
intron	complement(9740..9882)
	/gene="At4g14790"
	/number=7
exon	complement(9883..10162)
	/gene="At4g14790"
	/number=8
intron	complement(10163..10285)
	/gene="At4g14790"
	/number=8
exon	complement(10286..10349)
	/gene="At4g14790"
	/number=9
gene	10918..12578
	/gene="At4g14800"
exon	10918..11075
	/gene="At4g14800"
	/number=1
CDS	join(10918..11075,11167..11290,12261..12578)
	/gene="At4g14800"
	/note="strong similarity to proteasome chain Hsc7-I protein, human contains EST gb:AI993573.1, Y09359, R65485, T22127, F13874"
	/codon_start=1
	/product="proteasome chain protein"
	/protein_id="CA878522.1"
	/db_xref="GI:7268226"
	/translation="MECVFGVNGFPAIVADPSAVHSIILKKKEDKIMTLDHSKLY AASGEPDRVQFTFYVOKNVSLYGFRNGIFLSTAAANFTRGELATRLRNPPSVNLT MAGSDKAGASLVYIDYIATLHKYDKAGAFGSGTFSITMDRHRSDMSVEALELVD KCLLEIRSRVIAAPPNPFVIRIVDKGAREVGWRISTADA"
	11076..11166
	/gene="At4g14800"
	/number=1
exon	11167..11290
	/gene="At4g14800"
	/number=2
intron	11291..12260
	/gene="At4g14800"
	/number=2
exon	12261..12578
	/gene="At4g14800"
	/number=3
exon	12928..13042
	/gene="At4g14810"
	/number=1
gene	12928..15983
	/gene="At4g14810"
	join(12928..13042,13394..13501,13559..13702,15165..15253,15348..15365,15846..15983)
	/gene="At4g14810"
	/note="similarity to Fun12p, Saccharomyces cerevisiae contains EST gb:AA042489, AI998638.1, T88290"
	/codon_start=1
	/product="hypothetical protein"
	/protein_id="CA878523.1"

	repeat_region	/db_xref="GI:7268227"
	Intron	/translation="MTIVDGNOKFLILRESLEKTCRNGGGGGAARVAVSDLRKR OSRSIORESKHRLPEKTTETPVSCARNMGSDYKKAREHDELLFTSQNRPRRRCLL RKRGGGGGEGEDREYEDENAVSITLRTPPKPTQIKPLINSNHHITMNSDOSNFNOYV DGSGSQGLINVENYQALAPKACHVEPPAPASCHSKINSLSMH"
	exon	13001..13024 13043..13393 /gene="AT4g14810" /number=1
	Intron	13394..13501 /gene="AT4g14810" /number=2
	exon	13502..13558 /gene="AT4g14810" /number=2
	Intron	13559..13702 /gene="AT4g14810" /number=3
	exon	13703..15164 /gene="AT4g14810" /number=3
	Intron	15165..15253 /gene="AT4g14810" /number=4
	exon	15254..15347 /gene="AT4g14810" /number=4
	Intron	15348..15365 /gene="AT4g14810" /number=5
	exon	15366..15845 /gene="AT4g14810" /number=5
	exon	15846..15983 /gene="AT4g14810" /number=6
	gene	18314..20558 /gene="AT4g14820" complement(18314..19925)
	exon	19925..20002 /gene="AT4g14820" /number=1
	gene	complement(join(18314..19925,20002..20558)) /gene="AT4g14820"
	CDS	complement(join(18314..19925,20002..20558)) /gene="AT4g14820" /note="similarity to 130 KD LEUCINE-RICH PROTEIN (LRP 130) (GP130) L130_HUMAN Contains Cytochrome c family heme-binding site signature AA682-687;Prokaryotic membrane lipoprotein Lipid
Query Match		4.1% Score 44.4; DB 8; Length 197405;
Best Local Similarity		49.6%; Pred. No. 3.7;
Matches 114; Conservative	0; Mismatches 116; Indels	0; Gaps 0;
Oy	303 gataatcggaatgcctagacacgcgtctaagaatgatcctctgfacacatttcttaaa	362
Db	82097 GTCTTGTAGTTCACAAAAAAGTTTTTGTAATAATTACTCTTTGCAAGTTGTACGA	82156
Oy	363 gccataatagcaatgataatagaaaagtggagctgccaagtccggaagacgtataatatt	422
Db	82157 ACGTAAAGAAAAAAAAGAAGCAAACTAGAAATTAGTAGACCACCAATTAAGCTAGGT	82216
Oy	423 taaacgatcctgcacaaaagaatgtaaaggcgcaaaccaataattgycgagtgtaattcc	482
Db	82217 TTACCAAGCCCGTGTATAGAAATTAAGGCCCAAATTTAAAAATGGGCTTAACCTAAGGC	82276
Oy	483 tgatgtagcatcaatcaaatctttaatcaatcaataagaatcagtcctt	532
Db	82277 CCAATACTGAAGAACTAGGTTCTCTCCGATCATTTTCTGCAATTACTATT	82326
RESULT 14		
ATPCA2		

LOCUS	202860 bp	DNA	linear	PLN 29-JUN-1999
DEFINITION	Arabidopsis thaliana	DNA chromosome 4,	ESSA I FCA contig	fragment
ACCESSION	No. 2			
VERSION	297337			
KEYWORDS	297337.2	GI:5302774		
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucotids II; Brassicales; Brassicaceae; Arabidopsis.			
AUTHORS	1 (bases 1 to 202860) Bevan, M., Stekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N., Kreis, M., Kavanagh, T., Entian, K. D., Rieger, M., James, R., Pridmore, P., Hatzopoulos, P., Obermayer, B., Dnestrov, A., Jones, J., Palme, K., Ansoy, W., Delseny, M., Bancroft, I., Mewes, H. W., Schueller, C. and Chalhatzis, N.			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 202860)			
AUTHORS	EU Arabidopsis sequencing project.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-JUN-1999) MIPS, the Max-Planck-Institut fuer Biochemie, am Kioferspitz 10a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk			
COMMENT	On Jun 30, 1999 this sequence version replaced gi:2244829. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with AFCA3 at the 5' end and an overlap with AFCA3 at the 3' end.			
FEATURES	Location/Qualifiers			
SOURCE	1. 202860			
	/organism="Arabidopsis thaliana"			
	/variety="Columbia"			
	/db_xref="taxon:3702"			
	/chromosome="4"			
	1. 70917			
	/organism="Arabidopsis thaliana"			
	/db_xref="taxon:3702"			
	/clone="BAC TBM02D4"			
	13673..13975			
	/gene="dl3405w"			
	13673..13975			
	/gene="dl3405w"			
	/number=1			
	13673..13975			
	/gene="dl3405w"			
	/note="weak similarity to N-methyl-D-aspartate receptor glutamate-binding chain, rat"			
	/product="hypothetical protein"			
	/codon_start=1			
	/protein_id="CAI0252.1"			
	/db_xref="GI:2244830"			
	/db_xref="SPTREMBL:O23329"			
	/translation="MAKSDLEGGGNLYFGMKSSSLRNAIFRLKLYSIISLQILVTVGVAIVFVRPIEFETETHRGLAVFVILLPLRLRVSFHILFHSFGSIVIA"			
	complement(join(15633..15841,16271..16328,16682..16731,16815..17057,17098..17157,17295..17739,18223..18333,18701..19012))			
	/gene="dl3410c"			
	complement(15633..15841)			
	/gene="dl3410c"			
	/number=1			
	15633..19012			
	/gene="dl3410c"			
	complement(join(15633..15841,16271..16328,16682..16731,16815..17057,17098..17157,17295..17739,18223..18333,18701..19012))			
	/gene="dl3410c"			
	/note="weak similarity to DNA-binding protein"			

Db 136823 CCAATAGTGAACCTAGTTCCTCTCCGCATCCATTTTCGATTTACTATT 136872

RESULT 15

BC016000/c 1182 bp mRNA linear PRI 24-OCT-2001
 DEFINITION Homo sapiens, clone IMAGE:433588, mRNA.
 ACCESSION BC016000
 VERSION BC016000.1 GI:16359070
 KEYWORDS

SOURCE

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1182)
 Strausberg, R.
 Direct Submission
 Submitted (22-OCT-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbio.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Kelleman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 38 Row: d Column: 13
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES source

Location/Qualifiers
 1..1182
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:433588"
 /tissue_type="Skin, melanotic melanoma, high MDR."
 /clone_lib="NIH_MGC_62"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"

BASE COUNT ~ 357 a 196 c 195 g 434 t
 ORIGIN

Query Match 4.1%; Score 44; DB 9; Length 1182;
 Best Local Similarity 52.7%; Pred. No. 7.2;
 Matches 118; Conservative 0; Mismatches 105; Indels 1; Gaps 1;

0Y 18 aacttaacaaactcaccataaactaaactaaactaatatataatgaataaactcttttt 77
 1 11 111 11 11 111 11 111 111 111 111 111 111 111 111 111 111
 Db 289 ATCTATCCCAAGAACACGGGGTTTATTACATATGCAATCTATATGGGTTT 230
 1 11 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 0Y 78 atcaatgatggccttgatgccttcttaactttatctcgtgcccgaattccaacgggaa 137
 1 11 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 229 TTTAAAGGTGGTGGGATGTTTACCCTATGTTGTTTATTGGATGCCCTAAGGCA 170
 1 11 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 0Y 138 tcaggaaatcaacgaaagaaacta-agtgaatgatcttaagctgaagatttcg 196
 1 11 11 11 11 111 111 111 111 111 111 111 111 111 111 111
 Db 169 TATTAAGCCGATATGCAATTACTACTGATGATTAATGTTCTTAATAACTGTGA 110
 1 11 11 11 11 111 111 111 111 111 111 111 111 111 111 111
 0Y 197 gaagaactgtaccggtaggatagaacgaagaacggaatgata 240
 1 11 11 11 11 111 111 111 111 111 111 111 111 111 111 111
 Db 109 AATAGAAAGAGGAAAGAACGACCAAGCCCAAGAACTCTTAAGAGA 66

Search completed: June 27, 2002, 20:18:48
 Job time: 9856 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 20:21:50 ; Search time 362.26 Seconds
(Without alignments)
5118.610 Million cell updates/sec

Title: US-09-727-769a-7

Perfect score: 1080
Sequence: 1 agttaaataaccaaccaac.....atcagcagtgctgtatatac 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: N_Geneseq_032802.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
25: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1080	100.0	1080	22	AAF90281
2	555	51.4	555	22	AAF90280
3	527.4	48.8	1080	21	AAZ49495
4	344.2	31.9	555	21	AAZ49494
5	45.4	4.2	465	21	AAC94109
6	45.4	3.9	7319	24	ABL34045
7	42.4	3.9	17131	24	ABL33053
8	41.8	3.9	1857	21	AAC65441
9	41.6	3.9	5987	24	ABL33563

C	10	41.4	3.8	5884	24	ABL34165	Human immune syste
C	11	40.8	3.8	6741	21	AAA10595	Gene encoding a su
C	12	39.6	3.7	19659	24	ABL32767	Human immune syste
C	13	39.4	3.6	19070	22	AAK86208	Human immune/haema
C	14	39.3	3.6	6292	22	AA546735	Tumour suppressor
C	15	38.8	3.6	1565	23	ABL07153	Drosophila melanog
C	16	38.8	3.6	2700	23	ABL11892	Drosophila melanog
C	17	38.8	3.6	3682	23	ABL07152	Drosophila melanog
C	18	38.8	3.6	5287	24	ABL33493	Human immune syste
C	19	38.6	3.6	6602	22	AAF97874	Human neuroblastom
C	20	38.4	3.6	5121	20	AA202070	Borrelia burgdorfe
C	21	38.4	3.6	8605	23	ABL32535	Human immune syste
C	22	38.2	3.5	2697	24	ABL13982	Drosophila melanog
C	23	38.2	3.5	2854	23	ABL13954	Drosophila melanog
C	24	38.2	3.5	8103	24	ABL33661	Human immune syste
C	25	38.2	3.5	17293	24	AA561057	Human gene regulat
C	26	37.8	3.5	4590	22	AAH24065	Yeast AOD9604-asso
C	27	37.8	3.5	8513	22	AA545354	Chemically pretrea
C	28	37.8	3.5	8513	22	AA546367	Tumour suppressor
C	29	37.6	3.5	6071	24	ABL32325	Human immune syste
C	30	37.6	3.5	6071	24	AA561076	Human gene regulat
C	31	37.6	3.5	8093	22	AA546435	Tumour suppressor
C	32	37.6	3.5	16688	24	ABL32320	Human immune syste
C	33	37.4	3.5	5183	22	AA545506	Chemically pretrea
C	34	37.4	3.5	5183	24	ABL33349	Human immune syste
C	35	37.4	3.5	8700	24	ABL33002	Human immune syste
C	36	37.4	3.5	13814	24	ABL33193	Human immune syste
C	37	37.4	3.5	16818	23	ABL18690	Drosophila melanog
C	38	37.2	3.4	5997	24	ABL38625	Human immune syste
C	39	37.2	3.4	6699	24	AA563320	Chemically pretrea
C	40	37.2	3.4	40862	24	ABL34073	Human immune syste
C	41	37	3.4	6012	22	AA546724	Tumour suppressor
C	42	37	3.4	12405	22	AA545330	Chemically regulat
C	43	37	3.4	12405	24	AA561143	Human gene regulat
C	44	37	3.4	15865	22	AAK90825	Human digestive sy
C	45	36.8	3.4	7498	24	ABL32257	Human immune syste

ALIGNMENTS

RESULT	1	
AAF90281	AAF90281 standard; DNA: 1080 BP.	
XX		
AC	AAF90281:	
XX		
DT	22-AUG-2001 (first entry)	
XX		
DE	Nucleotide sequence of a protein-deamidating enzyme.	
XX		
KW	Protein-deamidating enzyme; mineral absorption; food allergy; dough;	
KW	bakery; confectionery; ss.	
XX		
OS	Cryseobacterium sp.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	61..1023
FT		/*tag= a
FT		/product= "protein-deamidating enzyme"
PN	EP1106696-A1.	
XX		
PD	13-JUN-2001.	
XX		
PF	04-DEC-2000; 2000EP-0310768.	
XX		
PR	03-DEC-1999; 99JP-0345044.	
XX		
PA	(AMAN-) AMANO ENZYME INC.	
XX		
PI	Yamaguchi S;	
XX		

DR MPI: 2001-376907/40.
 DR P-PSDB: AAB84387.
 PT New enzyme for use in e.g. bakery has an ability to deamidate amido
 PT groups in a protein -
 XX
 XX Example 11; Page 23; 43pp; English.
 XX
 CC The present sequence encodes a protein-deamidating enzyme from
 CC Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
 CC groups in a protein by directly acting upon the amido groups without
 CC cutting peptide bonds and without cross-linking the protein. The enzyme
 CC thus reduces the mineral sensitivity of the protein and increases the
 CC soluble mineral content in the protein-mineral solution, improving the
 CC absorption of minerals in the human body. The enzyme reduces or removes
 CC the toxicity and allergenic property of the protein in a food
 CC e.g. allergy. The enzyme is useful for the improvement of dough in the
 CC field of bakery and confectionery e.g. for the production of crackers,
 CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean
 CC protein products, in various food articles e.g. meat or fish products
 CC and noodles; and for improving functionality of plant or animal protein.
 CC
 XX
 SQ Sequence 1080 BP; 371 A; 204 C; 203 G; 302 T; 0 other;

Query Match 100.0%; Score 1080; DB 22; Length 1080;
 Best Local Similarity 100.0%; Pred. No. 3.9e-281;
 Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agttaaaataaacaacacttaacaacaaacccatcaacttaacaaattcaattatt 60
 DB 1 agttaaaataaacaacacttaacaacaaacccatcaacttaacaaattcaattatt 60
 QY 61 atgaaataatctttttatcaatgatgagccttctgtaccgcttcaactttaattcctgt 120
 DB 61 atgaaataatctttttatcaatgatgagccttctgtaccgcttcaactttaattcctgt 120
 QY 121 gccgattccaacggaatcagaagaatcaacggaagaagaataagtgtaattct 180
 DB 121 gccgattccaacggaatcagaagaatcaacggaagaagaataagtgtaattct 180
 QY 181 aagcgtaaaagatttcggaagaagctgacggtgaggaatgagaagaacggaatgata 240
 DB 181 aagcgtaaaagatttcggaagaagctgacggtgaggaatgagaagaacggaatgata 240
 QY 241 aaggtgtcatattatgttaactcgcaactctctatgaattaaagccgcaaaagaatgag 300
 DB 241 aaggtgtcatattatgttaactcgcaactctctatgaattaaagccgcaaaagaatgag 300
 QY 301 cagtatatcggaatgcttagacagcgtgttaagaatgaattctctgtacacatttctta 360
 DB 301 cagtatatcggaatgcttagacagcgtgttaagaatgaattctctgtacacatttctta 360
 QY 361 aagcctaataagaatgaagaatgaagaagtgagctgcgaagccggaagaatgaat 420
 DB 361 aagcctaataagaatgaagaatgaagaagtgagctgcgaagccggaagaatgaat 420
 QY 421 tttaaaacgactcctgcacaaagaatgaagaagccaaacaaatattgacgagtgaatt 480
 DB 421 tttaaaacgactcctgcacaaagaatgaagaagccaaacaaatattgacgagtgaatt 480
 QY 481 cctgatgtagctacattaaattcttatcaatcaataaagaatcagcttcggttacc 540
 DB 481 cctgatgtagctacattaaattcttatcaatcaataaagaatcagcttcggttacc 540
 QY 541 tctaagggctctccatcagatcacatccatccctgtgacggagatgtatgaaaga 600
 DB 541 tctaagggctctccatcagatcacatccatccctgtgacggagatgtatgaaaga 600
 QY 601 gcccaataaatgagacaattcttaatgaacaacggtatgactgtgaaacaatttga 660
 DB 601 gcccaataaatgagacaattcttaatgaacaacggtatgactgtgaaacaatttga 660

QY 661 tacggaacctaagaagcatcaacaagaactgtctgtgtgagctgagctacacgttgca 720
 DB 661 tacggaacctaagaagcatcaacaagaactgtctgtgtgagctgagctacacgttgca 720
 QY 721 atattgtagctataaanaatgtcttcggagtaacggaanaaagaattatgatacctta 780
 DB 721 atattgtagctataaanaatgtcttcggagtaacggaanaaagaattatgatacctta 780
 QY 781 ctatttcaagcggtctcttaacagatagacagatgagaagaacgtgttaacacct 840
 DB 781 ctatttcaagcggtctcttaacagatagacagatgagaagaacgtgttaacacct 840
 QY 841 tgcgagctgcacgcgttccctcttactgtaactgcaggaatgttatacagaagt 900
 DB 841 tgcgagctgcacgcgttccctcttactgtaactgcaggaatgttatacagaagt 900
 QY 901 cctagaaattcactgtatgacaacaatcgtataccaactggttactgtactaa 960
 DB 901 cctagaaattcactgtatgacaacaatcgtataccaactggttactgtactaa 960
 QY 961 tttaactgttccggaatgtctccttcaactgcacggaatgtatccagctgtgatt 1020
 DB 961 tttaactgttccggaatgtctccttcaactgcacggaatgtatccagctgtgatt 1020
 QY 1021 taattaattgataatttacagacactgctcatattacagaatcagagtgctgtatat 1080
 DB 1021 taattaattgataatttacagacactgctcatattacagaatcagagtgctgtatat 1080

RESULT 2
 AAF90280
 ID AAF90280 standard; DNA; 555 BP.
 XX
 XX AAF90280;
 AC
 XX
 XX 22-AUG-2001 (first entry)
 DT
 XX
 XX
 DE Nucleotide sequence of a protein-deamidating enzyme.
 KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
 KW bakery; confectionery; ss.
 OS Cryseobacterium sp.
 PN EP1106696-A1.
 XX
 XX 13-JUN-2001.
 PD
 XX
 PF 04-DEC-2000; 2000BP-0310768.
 PR
 XX
 XX 03-DEC-1999; 99JP-0345044.
 PA (AMAN-) AMANO ENZYME INC.
 XX
 XX Yamaguchi S;
 PI
 XX
 DR MPI: 2001-376907/40.
 DR P-PSDB: AAB84386.
 PT New enzyme for use in e.g. bakery has an ability to deamidate amido
 PT groups in a protein -
 XX
 XX Claim 8; Page 22; 43pp; English.

The present sequence encodes a protein-deamidating enzyme from
 CC Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido
 CC groups in a protein by directly acting upon the amido groups without
 CC cutting peptide bonds and without cross-linking the protein. The enzyme
 CC thus reduces the mineral sensitivity of the protein and increases the
 CC soluble mineral content in the protein-mineral solution, improving the
 CC absorption of minerals in the human body. The enzyme reduces or removes
 CC the toxicity and allergenic property of the protein in a food
 CC e.g. allergy. The enzyme is useful for the improvement of dough in the

Db 314 tcaagaagcgtgttggaatgaaatcccggttcaatgtttctcttaaatgaatcaacaata 373
Oy 377 aaatagaagaagtgtgagtcgcgaagtcgcgaagacg taagatalttttaaacgactcga 436
Db 374 aaattgcagaagaagaaagacaacagatgatgacatccgttatctttaaatctgtatca 433
Oy 437 caaagaagataaagaagcaacaataattgtagagtgtaattcctgtatgtatgatacat 496
Db 434 acaagaagaagagaggtgaaagaacaagaagcagtcagttcttactgtatctgtgcaagc 493
Oy 497 taattcttatcaatcaataaagaatcagtcgttcgtaaccttgaagtcggtctctc 556
Db 494 tgaacagttattattaccagatcaaaaacacaggtctgcggaactcttcaacagactcttc 553
Oy 557 catgcatcaatcatgatatccctgttagacgagatgtatgtcaagaagccataaagatgagac 616
Db 554 ctgtgtcaacctcagatataccggttcgagtgatgtatgtcaaggtgctcaaaaatgagac 613
Oy 617 aaattctaataagaacagcgtatgactgtgaaataaacaattgtatacggaaactaagg 676
Db 614 aaatcctatgaaacgcgcggtatgactgtgaaagaagcagttcgtatatagttaacttgag 673
Oy 677 catcaacagaagaactgtgtgtggtgagctgagctacacagttgcaattgtgtaagtata 736
Db 674 ctctacagaagaacagtcgtgtatcatggtatatacagtaagaactcttgtaagcttca 733
Oy 737 aaatgtcttcggaagtaacgaaataaagaataatgatcctcatttcaacagcgtc 796
Db 734 aaatgtcttcggaagtaattgttgaataaagaatcatagatcttcattcttcacagcgtc 793
Oy 797 cgtgaacagatatagacatgtgagaacgcttggttaacacctctgcggtatcgtacatc 856
Db 794 cgtgaacagattcgtatgtagagcgtcgtatcacaacaacagcgtgagtcgtcgtcgt 853
Oy 857 tttccctatgtatgtatctacgcaggaatgttatattacagaagctagtaattcttacc 916
Db 854 tatcttcctacgcaataacagcagaataatgttattacagaagctcgcgtacgttacc 913
Oy 917 tgbatgacaacaatctgatacatataccaactgtgtactgataaatttcaactgtctccg 976
Db 914 tgbatgataacaactatgtaataccaatgtgtattataacaatattctcctcccttcag 973
Oy 977 gatgttctccttcaactgcacgcggaatgatatccgctgtgtatattat 1025
Db 974 gatgttctccttccccaagcacaagtgtagcaagcgtgtgtatatt 1022

RESULT 4
AAZ49494
ID AAZ49494 standard; DNA; 555 BP.
XX
AC AAZ49494;
XX
DT 04-APR-2000 (first entry)
XX
DE Chryseobacterium gleum protein-deamidating enzyme encoding DNA.
XX
KW Protein deamidating enzyme; soil bacterium; deamidation activity;
KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW protein engineering; surface hydrophobicity; toxicity; allergic;
KW mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.
XX
OS Chryseobacterium gleum 'JCM 2410'.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..555
FT /tag= a
FT /product= "Protein deamidating enzyme"
FT /function= "Deamidate amido groups in a protein"
FT /note= "Improves protein function"
XX
PN EP976829-A2.

PD 02-FEB-2000.
XX
XX PF 04-JUN-1999; 99EP-0304367.
XX
XX PR 04-JUN-1998; 98JP-0173940.
XX
XX PA (AMANO) AMANO PHARM KK.
XX
XX PI Yamaguchi S, Matsuura A;
XX
DR WPI: 2000-118552/11.
DR P-PSDB; AAI44582.
XX
XX
PT New enzyme for modifying and improving the function of proteins and/or
PT peptides has deamidating activity without causing cross linking -
PS Claim 8; Page 23; 57pp; English.
XX
XX The present sequence is the DNA encoding the protein-deamidating enzyme,
CC isolated from a new strain of soil bacterium, Chryseobacterium gleum
CC JCM 2410. The enzyme exerts the deamidation activity by directly acting
CC upon side chain amido groups in the protein in bonded state and
CC releasing side chain carboxyl groups and ammonia. It can deaminate high
CC molecular weight proteins, without cross linking and cleavage of peptide
CC bonds, to improve protein function. This sequence is used for protein
CC engineering, to cause an increase in surface hydrophobicity and improve
CC the function of a plant or animal protein. It can also be used to remove
CC or reduce toxicity or allergic property of proteins in food, decrease
CC mineral sensitivity of protein, to allow greater absorption into the body
CC and to solubilise calcium for use in drinks and mineral enhancing agents.
XX
SQ Sequence 555 BP; 159 A; 122 C; 112 G; 162 T; 0 other;
Query Match 31.9%; Score 344.2; DB 21; Length 555;
Best Local Similarity 76.7%; Pred. No. 6, 5e-83;
Matches 421; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
Oy 472 agtgttaattcctgtatgtatgaatcaatcaatcttcttatacatcaataaagaatcagttc 531
Db 7 agtgttattcctgtatgtatgaatcaatcaatcttcttatacatcaatcaataaagaatcagttc 66
Oy 532 tgcgtgtaccttaagcgtcctcacaatgatacatcaatcaatcaatcaatcaatcaatcaat 591
Db 67 tgcggaactcttaacagcattctctctgtatcaacttcaagatccggttgaagatgt 126
Oy 592 tatgcaagaagccataagaatgagacaataatcaataaagaacggtatgactgtgaaaaa 651
Db 127 tatgcaagaaggtccacaataaagaacaataatccattgaaacgcggtatgactgtgaaaa 186
Oy 652 caattgtatacgaagaacctaaagcatcaacagaagaactgtgtgtgctgtgagctac 711
Db 187 cagttcgtatagtatctgtgaggtcttcaaggaacatgctgtgtatcatgtggtat 246
Oy 712 caagttgcaatatgtgttaagctataaataatgcttcggaagtaacggaataaagaattat 771
Db 247 caagttgcaaatgtgttaagctcaaaaatgcttcaggaattgtgtgaaanaagaacata 306
Oy 772 gatcttcaactattttcaagcgttcctgtgaagaagaacagatgtaggaacgcttcgct 831
Db 307 gatcttcaactattcttccagcgttcctgtgaagaagaacagatgtaggaacgcttcgct 366
Oy 832 aacaccttgcgagatcgcgtatcctctcttattagtaatactgcagaagaatgttat 891
Db 367 aacacaaggtcgcgagatcgcgtatcctctcttattagtaatactgcagaagaatgttat 426
Oy 892 tacagaagtcctagtaattcctaactgtatgacaacaatctgatacaatacaactgtgt 951
Db 427 tacagaagtcgcgtcaggttcatctatgatagtaacaacaatgtagatacaaatgtgt 486
Oy 952 ctgactaaatttcaactgttcttcggaatgttctctcaccgcgcggaagtagtatccagc 1011
Db 487 ttaaacatatctcacccttccagagatgttctccttccccaagcacaagatgtatgaagc 546

Best Local Similarity 51.28; Pred. No. 0.021;

Sequence 7319 BP; 1946 A; 263 C; 1552 G; 3558 T; 0 other;

Query Match 4.2%; Score 45.4; DB 24; Length 7319;
 Best Local Similarity 46.7%; Pred. No. 0.058;
 Matches 179; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

5 aaaaataccaaacccaacttaacaaacacccaacttaacataacataatattatga 64
 6212 AAACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6153

65 aaaaatcttttcttaacatgacgcttgcgcttgcgcttgcgcttgcgcttgcg 124
 6152 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6093

125 attccaagcggaatcgaagaatcgaagaagaagaagaagaagaagaagaaga 184
 6092 A--CTACCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6036

185 tgaagaattcggaaagactgtacggtagatagacgaagaagaagaagaagaaga 244
 6035 ATAAATAACGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5976

245 tgcattatctgttaactgcgcaatctatgaataatgaagcgaagaagaagaaga 304
 5975 TATATATTAAACAAATTTTAAACAAAAACAAATTTTATATATCAAAAAACAACTACT 5916

305 atatcggaatgcttagacagcgctgttaagaatgaatcctcctgtaacattctta 364
 5915 AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5856

365 ctaataagcaatgaagaagaagaagaagaagaagaagaagaagaagaagaaga 387
 5855 CTAAATTAACCTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5833

RESULT 7
 ABL33053/C
 ID ABL33053 standard; DNA: 17131 BP.

ABL33053:
 26-MAR-2002 (first entry)

Human Immune system associated gene SEQ ID NO: 1026.

Human; immune system disease; cytosine methylation; antislutimatic;
 antiarteriosclerotic; antihaemic; cytosolic; neotropic;
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 gene; ds.

Homo sapiens.
 WO200200928-A2.
 03-JAN-2002.
 02-JUL-2001; 2001WO-EP07537.
 30-JUN-2000; 2000DE-1032529.
 01-SEP-2000; 2000DE-1043826.
 (EPIC-) EPIGENOMICS AG.
 Olek A, Piepenbrock C, Berlin K;
 WPI; 2002-130909/17.
 Nucleic acid comprising fragment of chemically modified gene, useful
 for diagnosis and treatment of diseases associated with abnormal
 cytosine methylation

PS Claim 1; SEQ ID NO 1026; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX

Sequence 17131 BP; 5134 A; 253 C; 3328 G; 8413 T; 3 other:

Query Match 3.9%; Score 42.4; DB 24; Length 17131;
 Best Local Similarity 52.9%; Pred. No. 0.51;
 Matches 91; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

23 aacaaataccaccataaactaaataacataatgaagaatctttttatcaa 82
 1443 AACTATAACCAATTAACCAATTAACCAATTTTATTAATTAATTAATTAATTA 1384

83 tgatggccttctgacgcgttcaacttcaatccctgcgctgcgaatccgaacggaatcag 142
 1383 AATATACCTATACCCACTTATTTACATTAATTAATTAATTAATTAATTAATTA 1324

143 aaatcaacggaagaagaagaagaagaagaagaagaagaagaagaagaagaatt 194
 1323 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 1272

RESULT 8
 AAC65441
 ID AAC65441 standard; DNA: 1857 BP.

AAC65441:
 14-FEB-2001 (first entry)

Staphylococcus warneri Swai methylase gene.

Staphylococcus warneri; Swai restriction-modification gene;
 restriction endonuclease; Swai methylase; ds.

Staphylococcus warneri.
 EP1048731-A2.
 02-NOV-2000.
 26-APR-2000; 2000EP-0303507.
 27-APR-1999; 99US-0299378.
 (NEW) NEW ENGLAND BIOLABS INC.
 Kong H, Higgins LS, Dalton MA;
 WPI; 2000-681207/67.
 P-PSDB; AAB28685.
 New method for cloning and producing the Swai restriction endonuclease
 from Staphylococcus warneri which can be produced in abundance from
 Escherichia coli -
 Disclosure; Fig 3; 18pp; English.

The present sequence encodes the Staphylococcus warneri Swai
 methylase of the Swai restriction-modification system. Swai restriction
 endonucleases are used for characterising genes in the laboratory. The
 bacterial strains used permit better and simplified purification methods.
 Sequence 1857 BP; 793 A; 203 C; 220 G; 641 T; 0 other:

[illegible]

xx	Sequence	5987 BP; 1928 A; 96 C; 1209 G; 2754 T; 0 other;
sq		
	Query Match	3.9%; Score 41.6; DB 24; Length 5987;
	Best Local Similarity	46.5%; Pred. No. 0.57;
	Matches 134; Conservative	0; Mismatches 154; Indels 0; Gaps 0;
Qy	6 aaataaccaaccaacttaacaaaacacacactaaactaactaactatataatgaa	65
Db	534 AAATATCATCTATTATATTAACAAACAAATCTATCATATATATATTAATATTATTA	475
Qy	66 aaatcttttttatacaagatgagccttctgtgacgcgtctaactttaactctctg	125
Db	474 CCTTTTATTTAATTAATTAACAAACAAATTTTATTTAATATCTAATCAATTC	415
Qy	126 ttccaacggaatcaggaatcaacggaagaagaaactaagtgtaaatgactcaagct	185
Db	414 AAAAAACCCGAATTTATTATTTTAAATTAACAAACAAATATCTTAAAAATTAAC	355
Qy	186 gaaagatttcggaanaagacttaccggttaaggatagaagaaacggaatgataaagt	245
Db	354 TAAATTTTCCCTAATATATTATTAATATACTAAATCCATTTTAAATTTAAAAA	295
Qy	246 gtcaattatgttaactgcgaactctatagaactaagccgaccaaga	293
Db	294 ACTACTTATTTAACTCTTAAAAATTAACAAATTTTAACTCTTAAAAA	247

RESULT	10
ABL34165/c	
ID	ABL34165 standard; DNA; 5884 BP.
XX	
AC	ABL34165;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 2138.
XX	
KW	Human; immune system disease; cytosine methylation; antiasthmatic;
KW	antiartherosclerotic; antianaemic; cytosolic; neutrotropic;
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW	antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIC-) EPICENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI: 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation
XX	
XX	Claim 1; SEQ ID NO 2138; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences

PR 20-OCT-2000: 2000US-0241809.
 PR 20-OCT-2000: 2000US-0241826.
 PR 01-NOV-2000: 2000US-0244617.
 PR 08-NOV-2000: 2000US-0246474.
 PR 08-NOV-2000: 2000US-0246475.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246478.
 PR 08-NOV-2000: 2000US-0246523.
 PR 08-NOV-2000: 2000US-0246524.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246527.
 PR 08-NOV-2000: 2000US-0246528.
 PR 08-NOV-2000: 2000US-0246532.
 PR 08-NOV-2000: 2000US-0246609.
 PR 08-NOV-2000: 2000US-0246610.
 PR 08-NOV-2000: 2000US-0246611.
 PR 08-NOV-2000: 2000US-0246613.
 PR 17-NOV-2000: 2000US-0249207.
 PR 17-NOV-2000: 2000US-0249208.
 PR 17-NOV-2000: 2000US-0249209.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249312.
 PR 17-NOV-2000: 2000US-0249213.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249215.
 PR 17-NOV-2000: 2000US-0249216.
 PR 17-NOV-2000: 2000US-0249217.
 PR 17-NOV-2000: 2000US-0249218.
 PR 17-NOV-2000: 2000US-0249244.
 PR 17-NOV-2000: 2000US-0249245.
 PR 17-NOV-2000: 2000US-0249264.
 PR 17-NOV-2000: 2000US-0249265.
 PR 17-NOV-2000: 2000US-0249297.
 PR 17-NOV-2000: 2000US-0249299.
 PR 17-NOV-2000: 2000US-0249300.
 PR 01-DEC-2000: 2000US-0250160.
 PR 01-DEC-2000: 2000US-0250391.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.
 PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251989.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM:
 XX WPI: 2001-483426/52.
 XX
 PT Nucleic acids encoding human Immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Disclosure: SEQ ID NO 41020: 3071pp + Sequence Listing: English.
 XX
 CC AAK54951 to AAK64702 encode the human Immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)

CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat Immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human Immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 CC
 XX
 SO Sequence 15070 BP: 5077 A; 4466 C; 4474 G; 5053 T; 0 other:

 Query Match 3.6%; Score 39.4; DB 22; Length 19070;
 Best Local Similarity 49.8%; Pred. No. 3.4;
 Matches 100; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

 QY 325 gctgttaagaatgaatcctcctgtacacatttctttaaagcctaatgcaatgaatga 384
 Db 18683 gatgtaataatgaatgaaggcaaatattttagatataaaaggaaggaggg 18742
 QY 385 aaagtggagctgcaagtcgcgaagacgtaagattttaaacgactcgcacaaagaa 444
 Db 18743 gagatgcttctgtcattctctgtgacagtttgttccctacaataaacaatga 18802
 QY 445 gtaaaaggcaacccaataattgctgagtgtaattcctgattgtagctacataaattct 504
 Db 18803 taaaagggaaccacacattcattgagttccagggggaagtgcattgataatga 18862
 QY 505 ttattcaatcaataaagaat 525
 Db 18863 taatggaattcaatgaagaat 18883

 RESULT 14
 AAS46735/C
 ID AAS46735 standard; DNA; 6292 BP.
 AC AAS46735;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #459.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CPG dinucleotide; single-nucleotide polymorphism; SNP;
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001: 2001WO-EP02955.
 XX
 PR 15-MAR-2000: 2000DE-1013847.
 PR 06-APR-2000: 2000DE-1019058.
 PR 07-APR-2000: 2000DE-1019173.
 PR 30-JUN-2000: 2000DE-1032529.
 PR 01-SEP-2000: 2000DE-1043826.
 XX
 PA (EPTG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer -
 XX
 PS Claim 1: SEQ ID NO 459; 27pp; English.

Db 990 gcacaaagtagagagctcccaa 1011

Search completed: June 27, 2002, 20:22:25
Job time: 6443 sec


```

Db 194 gtaagactgtccggtgaggaattgaaagaatggaataatttaaatctcattatg 253
QY 257 taactgcgaattctctatgaatgaacgcgaacaaagaatgagcatalatcgaaatg 316
    ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 254 ttacgtccgacgcgtctgaatctgcgcagacgaataaagaatgaggtatattccatga 313
QY 317 ttgacaggtcgtttaagaatgaatctctgttacacattcttcttaagcctaagaatg 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 tcagagaggtctgtggaatgaatccgccttcattgtttctcttaagaagcaaaccaata 373
QY 377 aagaatgaagaatgtagctgtgcgaagtcgcgaagcgaatgaatattttaaagatctga 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 aatctgcaaaagaaagaaagaaagacagatgataccgttaattttaaactctgttca 433
QY 437 caaaagaagtaaaagggcaaacacataaaattggcgagtgtaattctctgtatgatact 496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 434 acaagaagaagagagggtggaagaacaaagaagcagtcagtggtattctctgtctgcaagc 493
QY 497 taaattcttattcaatacaataagaatcagtcctgtcggttacctctacagcgctctcaac 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 tgaacagttattatcccgacatcaaaaacagcgttgcgaaactctacagacatctctc 553
QY 557 catgtacatctacagatctcctgtgacgagtggtatgaagaagccgaatgaatgagc 616
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 ctgtatccgctccagatataccggtgcaggtgatagcaagggctccaaatgagac 613
QY 617 aaatcttaatgaacaacgcgtatgactgtgaataaacaattgtatagcgaataaag 676
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 aatctctatgaaacgcgcgtactgactgtgaaaagagtcgtatattgttaattctgagag 673
QY 677 catcaacaggaactgtcgtgtggtgagcgtacacagctgtcaatatgtgtaagctata 736
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 674 cctctcaagaacatctgtgtgatacgtggtatatacagcgaacatttgytaagcttca 733
QY 737 aaaaatgctccggaagaacgaaaagaatattgatacttcaactttaaagagtc 796
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 734 aaaaatgctccaggaattgtctgaaaaaagaatacagatccctcattatctccaagagtc 793
QY 797 ctgtacacagatacagatgagaaacgctgtgttaacacctctctgcgagatctgacacg 856
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794 cgttaacagatctctgatagagagcgtgatacacaacaacagcgtacatcgctcgtcgt 853
QY 857 ttctctctatgtactatctgcaggaatgtttatcacgaagtccttaattatcttacc 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 854 tatctctcaccgcaatacagcagaagaatgttactacagaagtcgcgtcaagttctaac 913
QY 917 tgtatgacaacaatctgatacacaactgtgtactgactaacttcaacttgcagcttcg 976
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 914 tgtatgataacaactatgtgataccaattgtgtataaataatctcaatcccttcag 973
QY 977 gatgtctctcaccctgcagcagatgatacgaagctgtgatttaatt 1025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 974 gatgtctctcctccacagaccaaagtgtgaagcgtgtgatttaatt 1022

```

```

; TYPE: DNA
; ORGANISM: Chryseobacterium g1eum
; US-09-324-910-5

Query Match      31.9%: Score 344.2; DB 4; Length 555;
Best Local Similarity 76.7%: Pred. No. 9,1e-91;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 472 agtgaattcccgatgtgaactatcatatttaattcttttcaatcaataaagaatcagttct 531
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7 agtgtattcccgatcttgcgaacgctgcaagcttcaatttcaaccagatccaanaacagct 66
QY 532 tgcgtacactctacagcgctgcctctcaacatgcatcaatcaatctgttaagcagatgt 591
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67 tgcggaactctcaacagcatctctctctgtatcaacttcaagatccggttgcagagtg 126
QY 592 tatgcaagggcccaataagatgagacaatcttaatgaaacaagcgtatgactgtgaaaaa 651
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 tatgcaagggctcaacaataatgagacaaatctatgaaacgcggtatgactgtgaaaag 186
QY 652 caattgtatagcgaacctaagacatacaacaggaactgtgtgtggtgagctac 711
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 cagttcgtataggttaattcgtgaagcttctcaaggaacatgctgtatcatatggtat 246
QY 712 caggttgcataatgtgtaagctataaaaaatgctccggaatgaacggaataaagaattatt 771
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 247 caggtgaacatttgttgaagctcaaaaatgcttcaaggaattgttgaaaaaagaatcata 306
QY 772 gatcttcaatttctcaagcggtctctgttaacagatacagcatgagaaacgctgtcgt 831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 gatcttcaatttctccagcggtctctgttaacagattctgtgagagatgctgttacc 366
QY 832 aacaccttgcgagatctgacccgttctctcttgaactgaactgacgaagaattgttat 891
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 aacacaagctcggaatcgtcgtctgtatcttctcaagcaatacagcaggaatgtttac 426
QY 892 tacagaagtcctagtaattcttaactgtatgaaacaacatctgataccaactgtgta 951
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 taacagaagtcgcaggttcatcatctgatagtataacaactatgtgataccaactgtgta 486
QY 952 ctgactcaatttcaactgtcttccggaatgttctctctcaactcagcaggaatgtatccag 1011
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 ttaacaatatctcattcccttcaagatgttctctccacagcacaagtgtagaagc 546
QY 1012 tgtgattt 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 tgtgattt 555

```

```

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22133-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 555

```

APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300,6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 5.6%, Score 60; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 1.3e-07;
Matches 6; Conservative 231; Mismatches 141; Indels 0; Gaps 0;

Oy 126 ttccacggaggaatcgagaaatcaacgaaagaaactaagtgttaattgattcttaagct 185
Db 1441 TTGGTACRRR 1382
Oy 186 gaaagattcggaaagactgtaacgtaggagtagacgaaagaaacggaatgataaggt 245
Db 1381 RR 1322
Oy 246 gtcatcttgtaactgcgcaattctatgaatgaaccgcaagaagaatgagcagta 305
Db 1321 RR 1262
Oy 306 tatcggaatgcttagacgagcgttaagaatgaatctctcgtacacattctttaaagcc 365
Db 1261 RR 1202
Oy 366 taatagcaatgaatgaagaatgagctgcaagtcggaagacgtaagaatatttaa 425
Db 1201 RR 1142
Oy 426 aacgatacctgcacaaagaagtaaaaggcaaaccaataatgycgagtgtaattcctga 485
Db 1141 RR 1082
Oy 486 ttagctacattaaattc 503
Db 1081 RRRRRRRRRRRRRRRATC 1064

RESULT 4
US-09-299-378-3
Sequence 3, Application US/09299378
Patent No. 6245545
GENERAL INFORMATION:
APPLICANT: Higgins, Lauren S.
APPLICANT: Dalton, Michael A.
APPLICANT: Kong, Huimin
TITLE OF INVENTION: Method For Cloning And Producing The Swai Restriction
FILE REFERENCE: Swai
CURRENT APPLICATION NUMBER: US/09/299,378
CURRENT FILING DATE: 1999-04-27

NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1857
TYPE: DNA
ORGANISM: Staphylococcus warneri
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1854)
US-09-299-378-3

Query Match 3.9%, Score 41.8; DB 4; Length 1857;
Best Local Similarity 48.9%; Pred. No. 0.013;
Matches 112; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

Oy 302 agtatatcgaaatgcttagacagcgtgtaagaatctctcgtacacatttcttaa 361
Db 905 agcaacttgatatacatatttaaaagctttagtgatgtagtgagacaataatcatctt 964
Oy 362 agccataatgacaatgaatgaagaagtgagctgcaagtcggaagacgtaagatatt 421
Db 965 ctctttaaaaatgcatccaacaaagaagctatgtaagatcaaaagattgccaataa 1024
Oy 422 ttaaaagctctcgacaaagaagtaaaaggcaaaccaataatgycgagtgtaattc 481
Db 1025 ataacacattccaattatcaatgaatagatggtgcatgataaattatctatgattta 1084
Oy 482 ctgatgtagctacattaaattcttatccaatccaataaagaatcagtc 530
Db 1085 tttagcgaagaagaataatcaattttagagaagtttacacataatc 1133

RESULT 5
US-09-141-047-7/c
Sequence 7, Application US/09141047A
Patent No. 6043085
GENERAL INFORMATION:
APPLICANT: Walker, David H.
APPLICANT: Yu, Xue-Jie
TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
FILE REFERENCE: D6143
CURRENT APPLICATION NUMBER: US/09/141,047A
CURRENT FILING DATE: 1998-08-27
NUMBER OF SEQ ID NOS: 15
SEQ ID NO 7
LENGTH: 2489
TYPE: DNA
ORGANISM: Ehrlichia canis
FEATURE:
LOCATION: -340..2149
OTHER INFORMATION: Nucleotide sequence of gene encoding 120kDa
OTHER INFORMATION: Immunoreactive protein
US-09-141-047-7

Query Match 3.3%, Score 35.6; DB 3; Length 2489;
Best Local Similarity 45.9%; Pred. No. 0.95;
Matches 122; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

Oy 741 tgcctcgaggaagaaagaaatgatgatctctcattcaatcttaagcggtccgt 800
Db 1843 TTCTTCTTACTAGTTTACGATACTTTTCTTCAACTTCACTTGATGATGTTTACTACT 1784
Oy 801 aacagatacagcatgagaaacgcttgtaacacctcttcgagatcgtacatccgttc 860
Db 1783 ACCATCTACAGCAGGTTCGAATCTTCGCTTACTTACAGAGATTTTCTTACT 1724
Oy 861 ctctatgctaatctgcaggaatgtttatatacagaagctcctagtaattcttactcgt 920
Db 1723 AGTTTCAGATACCTTTTCTCCAACTTCACTTGATGAATGTTCTACACTACATCTACAGC 1664

QY 921 tgaacaactcgtataccaactgtgtactgactaaatttcactcttcggaatg 980
 Db 1663 aggttgcaatcttctgctttacttcagagattttctcttactgatttcagatac 1604
 QY 981 ttctctactcgacggagatg 1006
 Db 1603 tttttctccacttcactgtgatg 1578

RESULT 6

US-09-092-770-8/c

; Sequence 8, Application US/09092770

; Patent No. 5973119

; GENERAL INFORMATION:

; APPLICANT: Coats, Steven R.

; APPLICANT: Bass, Michael B.

; APPLICANT: Robinson, Murray O.

; TITLE OF INVENTION: No. 5973119el Cyclin E Genes and Proteins

; FILE REFERENCE: A-524

; CURRENT APPLICATION NUMBER: US/09/092,770

; CURRENT FILING DATE: 1997-06-05

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO: 8

; LENGTH: 1215

; TYPE: DNA

; ORGANISM: Human

US-09-092-770-8

Query Match 3.3%; Score 35.4; DB 2; Length 1215;
 Best Local Similarity 24.8%; Pred. No. 0.77;
 Matches 121; Conservative 88; Mismatches 277; Indels 1; Gaps 1;

QY 115 ttcctgcccattccaacgggaatcaggaatcaacgaagaagaactaagtgtaat 174
 Db 773 tcyttnarngcrtcnacynarnarrtnarccanswdatdntngtnacngrcan 714
 QY 175 gattcctaagctgaagatttcggaaga-cgtaccgtaggagatagacgaagaacg 233
 Db 713 arytcccaayttnarncgcyttnaradatnarytccatncknaradttrcytcnsmr 654
 QY 234 aatgataaagggtcattatgttaactcgcgaatctatgaatlaagccgacaaga 293
 Db 653 cangcncrcctnctnacrtaangcraaytctgnarrttingncrtradttytctnary 594
 QY 294 aatgacagatatacgcgaatcgttagacagcgttgaagaatgaatccctgtacat 353
 Db 593 ttnsmwgcdratranarnrnsngtdatnccdatnarytgnarcrtttrtttdatrtcy 534
 QY 354 ttctctaaagcctaataagcaatgaatgaagaagtcgacgtcgcgaagcgaagcgt 413
 Db 533 ttttgngtnarcatranackrtccraaraartcytngcnarrtaraangtncckrtcn 474
 QY 414 aagatctttaaagaagatccctgacaaagaagaagtcgaagcgaacaaataatggcgag 473
 Db 473 arngttrtanacytcrcaancycrcanarccartcnararnaratnsmwncakcaytngcy 414
 QY 474 tgaattcctgagttagctacattaatcttattcaatcaataaagaatcagtcgt 533
 Db 413 tcnarrtctsmwrtgnarnrnacnycrcraartgtrtcrtnagnackrnksmswtctt 354
 QY 534 cggtaacttaagcggtcctcaccatgcatcacaatccagatataccgttagagatgta 593
 Db 353 arcattttanarccanacnycrtttnsmwncanccocansmwarrtcngnarnngsnngcr 294
 QY 594 tgcgaaga 600
 Db 293 ttdattra 287

RESULT 7
 US-09-222-851-8/c

; Sequence 8, Application US/09222851
 ; Patent No. 6165753
 ; GENERAL INFORMATION:
 ; APPLICANT: Coats, Steven R.
 ; APPLICANT: Bass, Michael B.
 ; APPLICANT: Robinson, Murray O.
 ; TITLE OF INVENTION: No. 6165753el Cyclin E Genes and Proteins
 ; FILE REFERENCE: A-524
 ; CURRENT APPLICATION NUMBER: US/09/222,851
 ; CURRENT FILING DATE: 1998-12-30
 ; EARLIER APPLICATION NUMBER: 09/092,770
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO: 8
 ; LENGTH: 1215
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-222-851-8

Query Match 3.3%; Score 35.4; DB 4; Length 1215;
 Best Local Similarity 24.8%; Pred. No. 0.77;
 Matches 121; Conservative 88; Mismatches 277; Indels 1; Gaps 1;

QY 115 ttcctgcccattccaacgggaatcaggaatcaacgaagaagaagaactaagtgtaat 174
 Db 773 tcyttnarngcrtcnacynarnarrtnarccanswdatdntngtnacngrcan 714
 QY 175 gattcctaagctgaagatttcggaaga-cgtaccgtaggagatagacgaagaacg 233
 Db 713 arytcccaayttnarncgcyttnaradatnarytccatncknaradttrcytcnsmr 654
 QY 234 aatgataaagggtcattatgttaactcgcgaatctatgaatlaagccgacaaga 293
 Db 653 cangcncrcctnctnacrtaangcraaytctgnarrttingncrtradttytctnary 594
 QY 294 aatgacagatatacgcgaatcgttagacagcgttgaagaatgaatccctgtacat 353
 Db 593 ttnsmwgcdratranarnrnsngtdatnccdatnarytgnarcrtttrtttdatrtcy 534
 QY 354 ttctctaaagcctaataagcaatgaatgaagaagtcgacgtcgcgaagcgaagcgt 413
 Db 533 ttttgngtnarcatranackrtccraaraartcytngcnarrtaraangtncckrtcn 474
 QY 414 aagatctttaaagaagatccctgacaaagaagaagtcgaagcgaacaaataatggcgag 473
 Db 473 arngttrtanacytcrcaancycrcanarccartcnararnaratnsmwncakcaytngcy 414
 QY 474 tgaattcctgagttagctacattaatcttattcaatcaataaagaatcagtcgt 533
 Db 413 tcnarrtctsmwrtgnarnrnacnycrcraartgtrtcrtnagnackrnksmswtctt 354
 QY 534 cggtaacttaagcggtcctcaccatgcatcacaatccagatataccgttagagatgta 593
 Db 353 arcattttanarccanacnycrtttnsmwncanccocansmwarrtcngnarnngsnngcr 294
 QY 594 tgcgaaga 600
 Db 293 ttdattra 287

RESULT 8
 US-09-004-393b-3
 ; Sequence 3, Application US/09004393B
 ; Patent No. 6310271
 ; GENERAL INFORMATION:
 ; APPLICANT: Hanson D., Andrew
 ; APPLICANT: Rathinasabapathi, Bala
 ; APPLICANT: Burnet, Michael
 ; TITLE OF INVENTION: Polynucleotides Encoding Choline Monooxygenase and
 ; FILE REFERENCE: UF-162


```

Oy      474  tgiatctcgatgttagctacatcaatcttataccaataagaagatcgcttg 533
Db      410  tcmarrctnsmrrctnarrnacttgaaragrgytrrrcrrgnmactrncrrtytcttcttgn 351
Oy      534  cggtaacctacaggcgcgtccatacatgcatcacaatcagatataccgttagacagatgta 593
Db      350  arcarrtrrrtccanacactyctgnsnancgccanmsnarrctnsgnarrngnsmngrr 291
Oy      594  tgcaga 600
Db      290  ttdata 284

RESULT 11
US-08-417-492-1
Sequence 1, Application US/08417492
Patent No. 5750872
GENERAL INFORMATION:
APPLICANT: Bennett, Alan B
APPLICANT: Brummell, David A
APPLICANT: Grantz, Alexander A
TITLE OF INVENTION: Nucleic Acids Encoding Ascorbate Free
TITLE OF INVENTION: Radical Reductase and Their Uses
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Stewart Street Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,492
FILING DATE: 05-Apr-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Baslian, Kevin L
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 2307E-586US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1723 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: exon
LOCATION: 49..643
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: intron
LOCATION: 644..780
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: 781..1484
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: CDS
LOCATION: join(49..643, 781..1487)
US-08-417-492-1

```

Query Match	3.2%	Score 34.2;	DB 1;	Length 1723;
Best Local Similarity	46.4%;	Pred. No. 2;		
Matches	111;	Conservative	0;	Mismatches 128; Indels 0; Gaps 0;
Oy	159	aaactaagtgttaaatgattctaagcttgaagaatttcggaaagactgtacgcgttggagat	218	
Db	890	AAATGAGAGGTGAAGGAAGTCAAACTCAAAATATGCGACAGCTTTTGAAAGCGACATAGT	949	
Oy	219	agacgaagaanaacggaaatcgataaagtgctattatgtttaactgcgcgaattcatgaat	278	
Db	950	AGTGTATGAGTTCGGAGCAGCAAGCACTCAGCACTTATTCAAAAGGCCAAGTTGAAGAGGA	1009	
Oy	279	taagcgcgaccaaagaanaattgagcagtatatacggaaatgctttagacagcgtgttaagaatga	338	
Db	1010	GAAGGTTGAATTAAAGACAGATGCTTCTTCAAAACAACTGACCTGATGTATATGCTGT	1069	
Oy	339	atctctctacacatttcttctaanaagccctaatacgaaataaataaggaanaagtgagtcgtc	397	
Db	1070	GGGTATGTTGGCACTTTTTCCTTTAAAAATGTACACATGAGATTAGAACAGTTGAACATG	1128	

```

RESULT 12
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-13

```

	Query Match	3.2%	Score 34.2;	DB 2,	Length 3666;
	Best Local Similarity	47.4%;	Pred.	No.2.9;	
	Matches 102;	Conservative 0;	Mismatches 113;	Indels 0;	Gaps 0;
Qy	737 aaatgcttcggagtaacgaaanaagaattatgcactcctaactatttccaagcggtc	796			
Db	1013 AAGGTACGTGATGAACATACTGCTTTATGCACACGGTGAICGTCCTAAATTTCACCTGGTT	1072			
Qy	797 ctgtaacagatacacgaatygagaaaacgcttgccgttaacaacccttcgsgatcgtacccg	856			
Db	1073 ATGATATCTGGGGTGTAATACAATTTCTTTCAGTTGAAAGAACTAATCAAGTGCTTCAG	1132			
Qy	857 ttctccttatgctcaatactcgcaagaaatggtttatatcaagaagtcctgtaattccttac	916			
Db	1133 TTAATATATGGTGCCAAACCAACTTACCAAAAGTTACTATTAAAAAACCCATAAACTGGTAAAC	1192			
Qy	917 tgtatgacaacaatcctgatcaaatlaccacaactcygta	951			
Db	1193 CAGAAGCAAACAAAACATTTAATGTTGGTTTGTGA	1227			
RESULT 13					
US-08-682-517-14					
; Sequence 14, Application US/08682517					
; Patent NO. 5874267					

```

: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Expression of surface layer proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,517
: FILING DATE:
: CLASSIFICATION:
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3666 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3666
:
US-08-682-517-14

Query Match          3.2%; Score 34.2; DB 2; Length 3666;
Best Local Similarity 47.4%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 737 aaatgcttcgcgagtaacggaagaaatattatgacccctcaatttcaagcgctc 796
    || || || || || || || || || || || || || || || || || || || ||
Db 1013 AAGGTACTGATGAGTAAGTACTGCTTATGCACTGGATCGATGTTCAAAATTCACACTGGTT 1072

Qy 797 ctgtaacagatacgcagatgagaaacgcttcgttaacacctcttcgagatcgcacgcg 856
    || || || || || || || || || || || || || || || || || || || ||
Db 1073 ATGTATCTGGGGGTGATACAAATTCCTTCAGTTGAAGAACTAAGTACAGTGTCTCAG 1132

Qy 857 ttctccttaagctaatcctgcaggaagtgttattacagaagtcctagtaattcttacc 916
    || || || || || || || || || || || || || || || || || || || ||
Db 1133 TTAATATGTTGCAACAAACCTTACAAAGTTACTTATATAAACCTTAACCTGTAAAC 1192

Qy 917 tgalgacaacaatctgatcaataccaactgtgta 951
    || || || || || || || || || || || || || || || || || || || ||
Db 1193 CAGAAGCAAAACAAACATTTAATGTGTGTTTGTGA 1227

RESULT 14
US-08-682-517-7
: Sequence 7, Application US/08682517
: Patent No. 5874267
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Expression of surface layer proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,517
: FILING DATE:
: CLASSIFICATION:
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4197 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
US-08-682-517-7
```

```

Query Match          3.2%; Score 34.2; DB 2; Length 4197;
Best Local Similarity 47.4%; Pred. No. 3.1;
Matches 102; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 737 aaatgcttcgcgagtaacggaagaaatattatgacccctcaatttcaagcgctc 796
    || || || || || || || || || || || || || || || || || || || ||
Db 1197 AAGGTACTGATGAGTAAGTACTGCTTATGCACTGGATCGATGTTCAAAATTCACACTGGTT 1256

Qy 797 ctgtaacagatacgcagatgagaaacgcttcgttaacacctcttcgagatcgcacgcg 856
    || || || || || || || || || || || || || || || || || || || ||
Db 1257 ATGTATCTGGGGGTGATACAAATTCCTTCAGTTGAAGAACTAAGTACAGTGTCTCAG 1316

Qy 857 ttctccttaagctaatcctgcaggaagtgttattacagaagtcctagtaattcttacc 916
    || || || || || || || || || || || || || || || || || || || ||
Db 1317 TTAATATGTTGCAACAAACCTTACAAAGTTACTTATATAAACCTTAACCTGTAAAC 1376

Qy 917 tgalgacaacaatctgatcaataccaactgtgta 951
    || || || || || || || || || || || || || || || || || || || ||
Db 1377 CAGAAGCAAAACAAACATTTAATGTGTGTTTGTGA 1411

RESULT 15
US-08-682-517-8
: Sequence 8, Application US/08682517
: Patent No. 5874267
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Expression of surface layer proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,517
: FILING DATE:
: CLASSIFICATION:
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4197 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Bacillus sphaericus
: INDIVIDUAL ISOLATE: P-1
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 95..3850
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 185..3850
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 95..184
:
US-08-682-517-8

Query Match          3.2%; Score 34.2; DB 2; Length 4197;
Best Local Similarity 47.4%; Pred. No. 3.1;
Matches 102; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 737 aaatgcttcgcgagtaacggaagaaatattatgacccctcaatttcaagcgctc 796
    || || || || || || || || || || || || || || || || || || || ||
Db 1197 AAGGTACTGATGAGTAAGTACTGCTTATGCACTGGATCGATGTTCAAAATTCACACTGGTT 1256

Qy 797 ctgtaacagatacgcagatgagaaacgcttcgttaacacctcttcgagatcgcacgcg 856
    || || || || || || || || || || || || || || || || || || || ||
Db 1257 ATGTATCTGGGGGTGATACAAATTCCTTCAGTTGAAGAACTAAGTACAGTGTCTCAG 1316
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2002, 19:22:05 ; Search time 2534.16 Seconds
(without alignments)
5752.088 Million cell updates/sec

Title: US-09-727-769a-7
Perfect score: 1080
Sequence: 1 agttaaataacacacacac.....atcagcaggtgctgtatat 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.8	5.2	1101	12	CNS00396
2	51.2	4.7	1101	12	CNS0182P
3	50.2	4.6	893	12	AZ538298
4	50.2	4.6	918	12	BH162500
5	50	4.6	942	12	CNS0186S
6	48	4.4	769	12	CNS011MN
7	47.8	4.4	734	12	CNS010MP
8	46.6	4.3	878	12	CNS0187R
9	46	4.3	1007	12	CNS06X9S
10	46	4.3	1101	12	CNS00LJT
11	45.8	4.2	997	12	CNS005TE
12	45.6	4.2	965	12	CNS010PW
13	45.6	4.2	965	12	BH137450
14	45	4.2	581	12	CNS01UPZ
15	45	4.2	964	12	CNS07EBR
16	45	4.2	1118	12	CNS02PY4
17	44.4	4.1	781	12	CNS009DO

C 18	44.4	4.1	928	12	CNS00DKY	AL071865 Drosophila
C 19	44.2	4.1	773	12	BH534872	BH534872 BOHAG04TF
C 20	44.2	4.1	959	12	CNS00655	AL062806 Drosophila
C 21	43.8	4.1	1101	12	CNS016LT	AL106896 Drosophila
C 22	43.6	4.0	1101	12	CNS00HX9	AL073856 Drosophila
C 23	43.2	4.0	1101	12	CNS00Z15	AL097091 Drosophila
C 24	43	4.0	1190	12	CNS020NT	AL097091 Drosophila
C 25	42.8	4.0	1101	12	CNS003B4	AL064082 Tetradon
C 26	42.8	4.0	1101	12	CNS00FT2	AL078714 Drosophila
C 27	42.6	3.9	1101	12	CNS00EVL	AL069706 Drosophila
C 28	42.4	3.9	886	12	CNS075DR	AL429961 clone_BAO
C 29	42.4	3.9	930	12	AZ693378	AZ693378 ENTIE73TR
C 30	42.4	3.9	969	12	BH154338	BH154338 ENYQY44TF
C 31	42.4	3.9	988	12	CNS00ZHB	AL097673 Drosophila
C 32	42.4	3.9	1204	12	CNS016E2	AL106628 Drosophila
C 33	42.2	3.9	562	12	CNS016OR	AL107013 Drosophila
C 34	42.2	3.9	974	12	CNS00ITT	AL075432 Drosophila
C 35	42.2	3.9	991	12	CNS00ITD	AL074343 Drosophila
C 36	42.2	3.9	1101	12	CNS012JN	AL101645 Drosophila
C 37	42	3.9	760	12	CNS011DI	AL100128 Drosophila
C 38	42	3.9	827	12	CNS00EJ3	AL069854 Drosophila
C 39	42	3.9	911	12	BH163485	BH163485 ENTS90TF
C 40	41.8	3.9	617	12	CNS031EC	AL223473 Tetradon
C 41	41.8	3.9	888	12	AZ547775	AZ547775 ENTF64TR
C 42	41.8	3.9	938	12	CNS006TJ	AL065906 Drosophila
C 43	41.6	3.9	1201	12	CNS0167M	AL106396 Drosophila
C 44	41.6	3.9	963	9	AL565565	AL565565 AL565565
C 45	41.6	3.9	1101	12	CNS017V6	AL108540 Drosophila

ALIGNMENTS

RESULT 1
CNS00396/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TERC end of BAC #
DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL063921 GI:4941778
VERSION AL063921.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr

COMMENT

- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Oseguwa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

SOURCE
location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"

BASE COUNT	201 a	64 c	131 g	202 t	503 others
ORIGIN					

Query Match	5.2%;	Score 55.8;	DB 12;	Length 1101;
Best Local Similarity	14.6%;	Pred. No. 0.011;		
Matches 71;	Conservative 228;	Mismatches 186;	Indels 0;	Gaps 0;

Qy 48 aattacaattatcttgaaaaatccttttttcaatcatgatgccttbytgacgcctctaac 107
::: ::|::: :::: : : ::| :| : : : :
Db 1084 DDDMTKMTWTWKDRADDRWAGDADRWANDGAGTWTATATWWWWWWATNTWTWCKW 1025

Oy 108 ttcttaattccctgtgcgcgaattccaacggaatcaaggaaatacagyaagaagaaaaactaag 167
::|||: | : :: | : ||| | : ||| : : ::
Db 1024 wmatTAAKTDTAWTMWRTAWRADWAGRDGACGRDRDAATDADGAGRDCGRRKKRDKKDR 965

OY 168 tctataatgatctctaagctgaaagtattcgcgaagaactgtacccgltaggatatagacaaga 227
::: : | :: || : : | :::: : : : | ::::| :: : |
Db 964 DGDGDDKKGKKKA KA KAA K A N A T T K W M D W D C D M K W G A K D R R A D D D D G A G D K D D D G K G K 905

OY 228 aaacggaatgatataaggtcattcttctgaactgcgaattcatgaaattaagccgac 287
 :|::|::::: ::::|::| ::::|::|::: ::
 Db 904 DADDDTGGKDDDDDKWDMDKKAKGTWGDATWAAATDMMWWMGADAWMTWDAADDN 845

Oy 288 caagaaanaatgacgatatacgaatgccttagacagcgtcttgaagaatcatctctgt 347
|:::| : | : : : || : ::|| : :
Db 844 WADDKMDAMWKKNDAMAWMGARTADRRDWGDSAGKRGGARKRDRKRADDDKDAADDRD 785

```

Qy 348 acacatttcttaagccataagcatgaatagaaagtgtgcgaagtcgga 407
      |::|::: : : : : : : : : : : : : : : : : : : : :
Db 784 AATTTTWTTRDDDKWKATDTWRWAADTWRDRODDDRRAGTAGKKWRPRW 725

```

QY 408 agcgcgaagatattttaaacgcatctgcacaaaagaagttaaagygcacaaccaataat 467
:: ::|||::| :::: : ::::: :: |::|:
Db 724 KRDDTRWDADADDTTARDRRRGGDGACAGCKKTKGKRRRDRATWDTDAWWADAAM 665

OY 468 ggcgagtgtaattcctgtagtcacatlaattctttatlccaataaaatca 527
 :
Db 664 wTTTTDTDDDDDKRDRRRRKGAARRRRRTTARAADWWTWKAWDAKWMTTRADRWRDWA 605

Qy	528	gtcct	532
	:	! !:	
Db	604	DTWTD	600

RESULT 2
CNS0182P

LOCUS	SIZE	DATE	DEFINITION
NSU182P	1101 bp	DNA	linear
			GSS 26-JUL-2004
			Drosophila melanogaster genome survey sequence Sp6 end of BAC
			BACN37D10 of DrosBAC library from Drosophila melanogaster (fru
			fly), genomic survey sequence.

ACCESSION	AL108811
VERSION	AL108811.1
KEYWORDS	GI:5629115
SOURCE	GSS.
	fruit fly.

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae. Drosophila

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1101)	Genoscope.	Direct Submission	Submitted (7/2-TUE-1989)
			Genoscope - Centro Nacional de Secuenciación

BP 191 91006 EVRY cedex - FRANCE (E-mail : secrete@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of the European Genomix project (Genomix-2) and in collaboration with the European Genomix project (Genomix-2).

Submitted (23-09-1999) genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
-Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billard at CEPR (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton

```

and Genevieve Payan. It has been constructed in the vector
pBel0BAC11.
Location/Qualifiers
1. 1101
features: "pBel0BAC11's reference"
source

```

	/note="end : SP6"			
BASE COUNT	274	a	268	c
ORIGIN			128	g
			73	t
				358 others

Query Match	4.78;	Score 51.2;	DB 12;	Length 1101;
Best Local Similarity	19.48;	Pred. No. 0.13;		
Matches	72;	Conservative 150;	Mismatches 150;	Indels 0; Gaps 0;

Qy 23 aacaaaaactcaccattaaactacaatttacaattatataagaaaaactttttttatcaa 82
| : | : | | | | | | | : : : : : | : | :
Db 683 ARAA.RAAAAA.DAKRAAAAAA.AAAAAA.AADAGKRRKKKGDKGDKGKATKTAAMAAG 742

Oy	83	tgaagcgcttgtgcacgclctaactttaattcctgtgcgatcccaacygaatcagg	142
	:::::	:::::	:
Db	743	RKDMGTATAMTWDTWADTYMKAATDIDAKRAAGRRKKDKARKTARDGGRARTPRAWA	802

Oy	143	aaatcaacggaaggaanaactaagtgttaatgatctctaagctgaagaatttcggaaga	2022
	: : : : : : :	: : : : : : : :	: : :
Db	803	AGGARRARAGARRAARRAADDDWDMAAAAAAAAAAAWTTTRDRRWKWDMDMTRWD	8622

```
QY 203 cctgtaacggttaggtagatagacgaagaacggaatgataaagctgcattatgttaactg 2622
      : || ::|| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 DDTTAAAMWDARARARRRRRRRRRRRRRAARRAADDDTKDRNADATDITDKDTIKWTY 9222
```

[illegible]

QY	383	gaaaagtggagt	394
	:	:	:
	:	:	:
DB	1043	DTDEKDDRTTADK	1054

RESULT 3
A2538298

LOCUS	893 bp	DNA	linear	GSS 14-NOV
DEFINITION	ENTGB51LR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.			
ACCESSION	A2538298			

ORGANISM	Entamoeba histolytica
SOURCE	Entamoeba histolytica
KEYWORDS	GSS.
VERSION	AZ538238.1 GI:11143106

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 893)	Loftus, B., Van Aken, S. and Fraser, C.	Determination of clone end sequences from <i>Entamoeba histolytica</i>
		Eukaryota; Entamoebidae; Entamoeba.

JOURNAL COMMENT
HML:IMSS sheared DNA Library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the *Entamoeba histolytica* HM1:IMSS sheared
DNA library
seq primer: M13-Reverse

Class: shotgun
High quality sequence start: 25
High quality sequence stop: 829.
Location/Qualifiers

FEATURES

1. 893

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone.lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a

light size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT

364 a 98 c 87 g 344 t

ORIGIN

Query Match 4.6%; Score 50.2; DB 12; Length 893;

Best Local Similarity 48.7%; Pred. No. 0.21; Mismatches 143; Indels 0; Gaps 0;

Matches 136; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 3 ttaaataaccacaactaacaaaactcaccattacaactaacattatattat 62

Db 352 TTATAATACCAATTATATATATATATATATATATATATATATATAT 411

Qy 63 gaaatactttttatcaatgatgaccttggtagccgtcttaacttccgtgc 122

Db 412 TTCAATATATGATTTATCAATGAAGATTTATGCGAGTTTATGATTAATGA 471

Qy 123 cgattccaacgggaatcggaatcaacggaagaagaactaagtgaatgtttctaa 182

Db 472 AGCATATCATTTCAATTCACAACTATTCAGTAAACAACTGTTTAAATGATTA 531

Qy 183 gctgaagattcggaaagactgtacgcgttagagtagaagaaagaaatgataaa 242

Db 532 TGTTCACACCATTTAGTAAACATTAATGATTTGAATTTCTAAACACTCCAAATTTAA 591

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Qy 592 TGAACGAATTTTATTTATATACACAAATATATAAATGA 630

Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 15
High quality sequence stop: 710.
Location/Qualifiers

FEATURES

1. 918

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone.lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a

light size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 385 a 105 c 92 g 336 t

ORIGIN

Query Match 4.6%; Score 50.2; DB 12; Length 918;

Best Local Similarity 48.7%; Pred. No. 0.21; Mismatches 143; Indels 0; Gaps 0;

Matches 136; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 3 ttaaataaccacaactaacaaaactcaccattacaactaacattatattat 62

Db 167 TTATAATACCAATTATATATATATATATATATATATATATATATAT 226

Qy 63 gaaatactttttatcaatgatgaccttggtagccgtcttaacttccgtgc 122

Db 227 TTCAATATATGATTTATCAATGAAGATTTATGCGAGTTTATGATTAATGA 286

Qy 123 cgattccaacgggaatcggaatcaacggaagaagaactaagtgaatgtttctaa 182

Db 287 AGCATATCATTTCAATTCACAACTATTCAGTAAACAACTGTTTAAATGATTA 346

Qy 183 gctgaagattcggaaagactgtacgcgttagagtagaagaaagaaatgataaa 242

Db 347 TGTTCACACCATTTAGTAAACATTAATGATTTGAATTTCTAAACACTCCAAATTTAA 406

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Qy 243 ggtgtcattatgttaactgcgaattctatgaataaa 281

Db 407 TGAACGAATTTTATTTATATACACAAATATATAAATGA 445

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

ACCESSION	AL108993
VERSION	ALI08993.1 GI:5629297
KEYWORDS	GSS.
SOURCE	Fruit fly. <i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachytera;
ORGANISM	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. Genoscope.
AUTHORS	I (bases 1 to 878)
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with The European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MC Project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.
FEATURES	Location/Qualifiers source 1..878 /organism="Drosophila melanogaster" /plasmid="pbelOBAC11" /db_xref="taxon:7227" /clone_lib="DrosBAC" /clone="BACN04E04" /note="end : SP6"
BASE COUNT	279 a 132 c 120 g 182 t 165 others
ORIGIN	
Query Match	4.3% Score 46.6 DB 12 Length 878;
Best Local Similarity	25.2%; Pred. No.1.5;
Matches	76 Conservative 88 Mismatches 137 Indels 0 Gaps 0;
OY	227 aaacggaatatabtaaggtgcattcatcgttactcgcaatctatatgaatgaaggcca 286
Db	41 AAAAAMMAATCAAAWRTTGTGAAAATAAAAAAAMMNCACAATAAAAAAAAAAMVANAAMAW 100
OY	287 ccagaagaatgatgcagtatacggaaatgcttagcacagtcgtttaagaatgaatccctc 346
Db	101 MMMAAAAAAAMMMARCAWMAGCWWGGGAWWWMMMMMARMMMAMWAAMAAAAAAMMWCTT 160
OY	347 taaccattttctaaggctaatgcgaatgaatacagaanaagtgagtcgtcaaagtcg 406
Db	161 TGCMCTGAMFTTGTTTRTWTAAMMMRTTRAAMMMMAAMMMMAAMMMWARCGMMA 220
OY	407 aagacgtatagattattaacagctacctgcacaagaagaataaagggccaaccaataat 466
Db	221 AAARAAMWAMMRRCRWTCRGMAMAAMMAAAAAAAAAAMMAAMMAAACRAMMMWMAAMCR 280
OY	467 ttgcgagltgaattctcctgatgtagctacataaatctttcataatcaatalaagaact 526
Db	281 AAMWGAMRAABMTBGAMWMBAAMAAAVRARBMARWAMWMAAMMATWTACAWAAMA 340
OY	527 a 527
Db	341 A 341
RESULT	9
CNSO6X9S	
LOCUS	CNSO6X9S 1007 bp DNA linear GSS 06-Jul-2001
DEFINITION	T3 end of clone AXOA03PF08 of library AXOA from strain CBS 7064
ACCESSION	AL419462
VERSION	AL419462.1 GI:12202640
KEYWORDS	GSS.
SOURCE	Pichia farinosa.
ORGANISM	Pichia farinosa

[illegible][illegible]

OY	245	tgcattatgcttaactagtcggcaatctcgttgaataacgccgacaagaataatgtagcgt	304
Db	333	AWAATAAATAAMTWTMTWMAAAWMAAWMACAAAAMAAAAAAGAAAAA	392
OY	305	atatcgaaatgcttgacgaaggctgttaaagtcatccctctgtcacacctttctaagc	364
Db	393	AATTAATAAAMATTCTTGCAWMAATTTAAATAAMMATTAATAAAARATTAATAATAA	452
OY	365	ctaatagcatagatgaataggaaaagtgtgagcttgtcaagtcgcgaagacgtaatatttta	424
Db	453	CWWAAAAAAAAAACAAAAAAMAAAAAGAAGAAAAAAGAAACMAAAAMATWTAAGMWACAATM	512
OY	425	aaac	428
Db	513	ATAC	516
RESULT_11			
CNS005TE/c			
LOCUS			
DEFINITION			
Drosophila melanogaster genome survey sequence TEn3 end of BAC #			
BACRICK22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
AL060767			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 997)			
Genoscope.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bdpac.med.buffalo.edu/drosophila_bac.htm .			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match 4.2%; Score 45.8; DB 12; Length 997; Best Local Similarity 17.4%; Pred No.2.3; Indels 0; Gaps 0; Matches 59; Conservative 129; Mismatches 151;			
OY	135	gaatcaggaataatcgcgaaaaagaaataagttlaaatagtatlttaagctgaagattt	194
Db	892	RAAFDARAARABARRRRRAKAKARARGARRRGRRRRGRAGRBRARAGGPRGRCGR	833
OY	195	cggaaagactgtaccggtgcggatgacgaagaaaaacggatcataaaggtgtcatttat	254

Db	832	GRGRARARARRAGARARARARARARARARRRRRAGARRRRRAGARRRRRGRRCR	773
Qy	255	gttaactcgcgaattcttgaataagcccgaccaaagaataagcagtatatcggaat	314
Db	772	GGRRGRRRRRARRARRARARAGARAAAAARRRRARRRGAAGAAARRRRRRRRRRRGCAGARAR	713
Qy	315	gcttaagaagcgtgtgaagaatgaatctccctgtacaacttcttctaagaagccataagca	374
Db	712	ARRRRARRARAGAGARRRRRRGRGAGARRRRGRGRRRCGARRRRRRRRRRRRRRR	653
Qy	375	tgaataagaagaagtgtgagtcgtcgaagtcgcgaagaagcagtaagataattlaaacagatcct	434
Db	652	RRRARCGAARRRRARRARARRRRRRGRGRRRRRRRAGRRRRRRRRRRRAGARRRRRR	593
Qy	435	gacaaaagaagtaaaaggcacaacccaataatgtgcgag	473
Db	592	AGRRRRRRRAGAGARR	554
RESULT 12			
CNS010Pw/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
SOURCE			
BASE COUNT			
ORIGIN			

RESULT	13
BH137450	
LOCUS	BH137450 965 bp DNA linear GSS 07-AUG-2001
DEFINITION	Entomobla histolytica Sheared DNA Entamoeba histolytica
ACCESSION	genomicC, BH137450
VERSION	BH137450.1 GI:15096511
KEYWORDS	GSS.
SOURCE	Entamoeba histolytica.
ORGANISM	Entamoeba histolytica
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.
AUTHORS	1 (Bases 1 to 965)
TITLE	Lofthus,B., Wang,Z., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica BH1:IMSS sheared DNA library (2001) Unpublished (2001)
JOURNAL	Contact: Brendan J Lofthus
COMMENT	

FEATURES
source

BASE COUNT
ORIGIN

4.28; Score 45.6; DB 12; Length 965;

[illegible]

LOCUS

ACCESSION
VERSION
KEYWORDS

AUTHORS

JOURNAL,

Source:

Location/Qualifiers
1. .581

```

/organism="Tetrarodon nigroviridis"
/db_xref="taxon:99883"
/clone="197M17"
/clone_lib="G"
/note="genoscope sequence ID : C0AC19/AG09SPL-end :
UC-Or1"

```


sequef@genoscope.cns.fr - Web : <http://www.genoscope.cns.fr>

Search completed: June 27, 2002, 19:22:12
Job time: 10620 sec

Search completed: June 27, 2002, 19:22:12
Job time: 10620 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:47:22 ; Search time 84.27 Seconds
(without alignments)
421.783 Million cell updates/sec

Title: US-09-727-769a-8
Perfect score: 1679
Sequence: 1 MNKFLSMMAFVTLTFNSC.....FSLLSGCSPPADVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A_Geneseq_032802.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1679	100.0	320	AA84387	Amino acid sequence
2	1292.5	77.0	319	AA14583	Chrysoobacterium g
3	991	59.0	185	AA84386	Amino acid sequence
4	874	52.1	185	AA14582	Chrysoobacterium g
5	116.5	6.9	884	AA83793	Bacillus cereus 10
6	112.5	6.7	852	AA81246	VIP1A(a) protein w
7	112.5	6.7	852	AA19516	Maize optimised-B.
8	112.5	6.7	852	AA846727	Maize optimised-B.
9	112.5	6.7	880	AA860224	Bacillus thuringie
10	112.5	6.7	881	AA159277	MIS toxin from B.
11	112.5	6.7	884	AA81239	B. cereus VIP1A(a)

12	112.5	6.7	884	18	AA19509
13	112.5	6.7	884	19	AA846712
14	112.5	6.7	1338	17	AA81247
15	112.5	6.7	1338	18	AA19520
16	112.5	6.7	1338	19	AA846731
17	112.5	6.7	1346	17	AA81245
18	112.5	6.7	1346	18	AA19513
19	112.5	6.7	1346	18	AA846723
20	108	6.4	20	22	AA84385
21	99	5.9	594	21	AA852451
22	99	5.9	594	21	AA852479
23	96	5.7	20	22	AA84384
24	95	5.7	1541	11	AA807304
25	93	5.5	264	21	AA825396
26	93	5.5	292	21	AA825395
27	93	5.5	294	21	AA825394
28	93	5.5	1350	22	AA863963
29	92.5	5.5	456	18	AA86050
30	91.5	5.4	450	19	AA868762
31	91.5	5.4	615	15	AA863230
32	91.5	5.4	1086	18	AA86051
33	91	5.4	812	21	AA873356
34	91	5.4	848	21	AA88565
35	91	5.4	1185	13	AA82675
36	90	5.4	477	22	AA830815
37	89.5	5.3	610	15	AA863229
38	89.5	5.3	1338	14	AA841731
39	89.5	5.3	1598	18	AA830291
40	89.5	5.3	2500	21	AA818272
41	89	5.3	661	22	AA808463
42	88.5	5.3	512	19	AA81554
43	88.5	5.3	1252	16	AA860530
44	88	5.2	458	22	AA861099
45	88	5.2	761	22	AA858195

ALIGNMENTS

RESULT 1	
AA84387	standard; Protein; 320 AA.
AA84387	
22-AUG-2001	(first entry)
Protein-deamidating enzyme; mineral absorption; food allergy; dough; bakery; confectionery.	
Chrysoobacterium sp.	
Key	Location/Qualifiers
Region	1..135
	/note- "prepro region"
EP1106696-A1.	
13-JUN-2001.	
04-DEC-2000; 2000EP-0310768.	
03-DEC-1999; 99JP-0345044.	
(AMAN-) AMANO ENZYME INC.	
Yamauchi S;	
WPI; 2001-376907/40.	
N-PSDB; AAF90281.	

B. cereus VIP1A(a)
100 kDa VIP1A(a) t
VIP2A(a) and VIP1A
Maize optimised-B.
VIP2A(a) and VIP1A
B. cereus VIP1A(a)
VIP1A(a) and VIP1A
Internal peptide o
Arabidopsis thalia
Arabidopsis thalia
N-terminal peptide
IgA1 protease. Ha
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Hybrid Marek's dis
H. pylori GPO 109
Mycoplasma gallise
Hybrid Marek's dis
HTM clone 2076520
Human NCAM 140KD 1
Collagen binding p
Amino acid sequenc
Mycoplasma gallise
High molecular wei
Non-lysable Hemo
Plasmodium falcipa
Novel human diagno
Collagen binding p
B. sphaericus SLP.
Drosophila melanog
Drosophila melanog

PT New enzyme for use in e.g. bakery has an ability to deamidate amido
 groups in a protein
 XX
 PS Example 11; Page 24; 43pp; English.
 CC The present sequence represents a protein-deamidating enzyme from
 CC Cryoseobacterium sp. number 9670. The enzyme is able to deamidate amido
 CC groups in a protein by directly acting upon the amido groups without
 CC cutting peptide bonds and without cross-linking the protein. The enzyme
 CC thus reduces the mineral sensitivity of the protein and increases the
 CC soluble mineral content in the protein-mineral solution, improving the
 CC absorption of minerals in the human body. The enzyme reduces or removes
 CC the toxicity and allergenic property of the protein in a food
 CC e.g. allergy. The enzyme is useful for the improvement of dough in the
 CC field of bakery and confectionery e.g. for the production of crackers,
 CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean
 CC protein products, in various food articles e.g. meat or fish products
 CC and noodles; and for improving functionality of plant or animal protein.
 CC
 XX
 SQ Sequence 320 AA:

Query Match 100.0%; Score 1679; DB 22; Length 320;
 Best Local Similarity 100.0%; Pred. No. 5.5e-150; Indels 0; Gaps 0;
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNFLSMMAEYVLTFTNSCADSNGNOETNGEKLKSVNDKLDKFGKTVPGIDEENGMI 60
 DB 1 mknflsmmaefvltftnscadsngnqetngelkksvndskldfgktvpgideengmi 60
 QY 61 KSPFMTAOFYELKPTKEBOYIGMROAVKNESPVHFLKNSNRTGKVESASPEDVRY 120
 DB 61 kspfmtaofyelpkptkneyigmravknespvhflknsneigrkvesaspedvry 120
 QY 121 FETILTKEKGGOTNKLASYIPVATLNSLFNQIKNSCSTASSPCIFRYPVDCGYAR 180
 DB 121 fetiltkekgotnklaasvtpvatlnslfnqikngscstasspcitfrypvdcgyar 180
 QY 181 AHKMRQILMNNGYDCEKQFVYGNLKAAGTCCVAMSYPHVAIIIVSYKNASGVTEKRIIDPS 240
 DB 181 ahkmrgilmmngydekgfvygnlkaagtcvawshvaailvsvknaasgvtekrilids 240
 QY 241 LFSSGPTDTARNACVNTSCGSASVSYANTAGNYTSPNSSYLDNNLINTNCVLK 300
 DB 241 lfssgptdtarnacvntscgsasvssyantaagnvytspnsylydnlnlntncvlk 300
 QY 301 FSLISGSPSPADVSSCGF 320
 DB 301 fslisgspspadvsscgf 320

RESULT 2
 ID AAY44583 standard; Protein; 319 AA.
 XX AAY44583;
 AC
 XX
 DT 04-APR-2000 (first entry)
 XX
 DE Chryseobacterium gleum protein-deamidating prepro-enzyme.
 XX
 KM Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;
 KM amido group; carboxyl group; ammonia; cross linking; peptide bond;
 KM protein engineering; surface hydrophobicity; toxicity; allergic;
 KM mineral sensitivity; calcium; absorption; mineral enhancing agent.
 XX
 OS Chryseobacterium gleum 'JCM 2410'.
 XX
 FH Key
 FH Region 1..134
 FT /label= "Prepro-region
 FT /note= "Comprises the Pre and Pro region"
 FT Peptide 1..21

FT /label= Signal_peptide
 FT /note= "Corresponds to the Pre region"
 FT 22..134
 FT /note= "Corresponds to the Pro region"
 FT 135..319
 FT /label= Mature_protein-deamidating-enzyme
 FT /note= "Deamidates amido groups in a protein"
 FT
 FT EP976829-A2.
 PN 02-FEB-2000.
 XX
 XX 04-JUN-1999; 99EP-0304367.
 XX
 XX 04-JUN-1998; 98JP-0173940.
 XX
 PA (AMAN) AMANO PHARM KK.
 XX
 XX Yamaguchi S, Matsuura A;
 PI
 DR WPI: 2000-118552/11.
 DR N-PSDB: AA249495.
 XX
 PT New enzyme for modifying and improving the function of proteins and/or
 PT peptides has deamidating activity without causing cross linking -
 XX
 PS Example 26; Page 26; 57pp; English.

The present amino acid sequence is the protein-deamidating
 CC prepro-enzyme, isolated from a new strain of soil bacterium,
 CC Chryseobacterium gleum JCM 2410. This enzyme exerts the deamidation
 CC activity by directly acting upon side chain amido groups in the protein
 CC in bonded state and releasing side chain carboxyl groups and ammonia. It
 CC can deamidate high molecular weight proteins, without cross linking and
 CC cleavage of peptide bonds, to improve protein function. This sequence is
 CC used for protein engineering, to cause an increase in surface
 CC hydrophobicity and improve the function of a plant or animal protein. It
 CC can also be used to remove or reduce toxicity or allergic property of
 CC proteins in food, decrease mineral sensitivity of protein, to allow
 CC greater absorption into the body and to solubilise calcium for use in
 CC drinks and mineral enhancing agents.
 CC
 XX

SQ Sequence 319 AA:

Query Match 77.0%; Score 1292.5; DB 21; Length 319;
 Best Local Similarity 74.7%; Pred. No. 1.7e-113;
 Matches 239; Conservative 39; Mismatches 41; Indels 1; Gaps 1;

QY 1 MKNFLSMMAEYVLTFTNSCADSNGNOETNGEKLKSVNDKLDKFGKTVPGIDEENGMI 60
 DB 1 mknflsmmaefvltftnscadsngnqetngelkksvndskldfgktvpgideengmi 60
 QY 61 KSPFMTAOFYELKPTKEBOYIGMROAVKNESPVHFLKNSNRTGKVESASPEDVRY 120
 DB 61 kspfmtaofyelpkptkneyigmravknespvhflknsneigrkvesaspedvry 120
 QY 121 FETILTKEKGGOTNKLASYIPVATLNSLFNQIKNSCSTASSPCIFRYPVDCGYAR 180
 DB 121 fetiltkekgotnklaasvtpvatlnslfnqikngscstasspcitfrypvdcgyar 180
 QY 181 AHKMRQILMNNGYDCEKQFVYGNLKAAGTCCVAMSYPHVAIIIVSYKNASGVTEKRIIDPS 240
 DB 181 ahkmrgilmmngydekgfvygnlkaagtcvawshvaailvsvknaasgvtekrilids 240
 QY 241 LFSSGPTDTARNACVNTSCGSASVSYANTAGNYTSPNSSYLDNNLINTNCVLK 300
 DB 240 lfssgptdtarnacvntscgsasvssyantaagnvytspnsylydnlnlntncvlk 300
 QY 301 FSLISGSPSPADVSSCGF 320
 DB 301 fslisgspspadvsscgf 319

RESULT 3
AAB84386
ID AAB84386 standard; Protein: 185 AA.
XX
AC AAB84386;
XX
DT 22-AUG-2001 (first entry)
XX
DE Amino acid sequence of a protein-deamidating enzyme.
XX
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;
KW bakery; confectionery.
XX
OS Cryseobacterium sp.
XX
PN EP1106696-A1.
XX
PD 13-JUN-2001.
XX
PF 04-DEC-2000; 2000EP-0310768.
XX
PR 03-DEC-1999; 99JP-0345044.
XX
PA (AMAN-) AMANO ENZYME INC.
XX
PI Yamaguchi S;
XX
DR WPI: 2001-376907/40.
DR N-PSDB: AAF90280.
XX
PT New enzyme for use in e.g. bakery has an ability to deaminate amido
PT groups in a protein
XX
PS Claim 4; Page 22; 43pp; English.
XX
CC The present sequence represents a protein-deamidating enzyme from
CC Cryseobacterium sp. number 9670. The enzyme is able to deaminate amido
CC groups in a protein by directly acting upon the amido groups without
CC cutting peptide bonds and without cross-linking the protein. The enzyme
CC thus reduces the mineral sensitivity of the protein and increases the
CC soluble mineral content in the protein-mineral solution, improving the
CC absorption of minerals in the human body. The enzyme reduces or removes
CC the toxicity and allergenic property of the protein in a food
CC e.g. allergy. The enzyme is useful for the improvement of dough in the
CC field of bakery and confectionery e.g. for the production of crackers,
CC biscuits, cookies, pizza pies or crusts or pie; in producing soybean
CC protein products, in various food articles e.g. meat or fish products
CC and noodles; and for improving functionality of plant or animal protein.
XX
SQ Sequence 185 AA;

Query Match 59.0%; Score 991; DB 22; Length 185;
Best Local Similarity 100.0%; Pred. No. 2.2e-85;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 LASVLPVATLNSLNFQIKNOSCTSTASSPCITFRFPVDCGYARAHKROIILMNGYDC 195
DB 1 lasvlpvatlnslnfqikngscgtstasspcitfrfpvdcgyarahkroiilmnngydc 60
QY 196 EKOPFYGNLKAFTGTCVAMSYHVAIIIVSYKNASGVTEKRIIDPSLSSGPTDTAMRNA 255
DB 61 ekqfygnlkaftgctcvawshvaailivsyknasgvtekrliidpslssgptdtawrna 120
QY 256 CVNTSCGSASVSYANTAGNYYRSPNSLYDNNLINTNCVLTFRFSLSGCSPSPADPV 315
DB 121 cvntscgsasvsyantagnyyrspnslydnnlntncvltfrfslsgcspspadv 180
QY 316 SSCGF 320
DB 181 sscgf 185

RESULT 4
AA44582
ID AA44582 standard; Protein: 185 AA.
XX
AC AA44582;
XX
DT 04-APR-2000 (first entry)
XX
DE Chryseobacterium gleum protein-deamidating enzyme.
XX
KW Protein deamidating enzyme; soil bacterium; deamidation activity;
KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
KW protein engineering; surface hydrophobicity; toxicity; allergic;
KW mineral sensitivity; calcium; absorption; mineral enhancing agent.
XX
OS Chryseobacterium gleum 'JCM 2410'.
XX
PN EP976829-A2.
XX
PD 02-FEB-2000.
XX
PF 04-JUN-1999; 99EP-0304367.
XX
PR 04-JUN-1998; 98JP-0173940.
XX
PA (AMAN) AMANO PHARM KK.
XX
PI Yamaguchi S, Matsuura A;
XX
DR WPI: 2000-118552/11.
DR N-PSDB: AAZ49494.
XX
PT New enzyme for modifying and improving the function of proteins and/or
PT peptides has deamidating activity without causing cross linking
XX
PS Claim 5; Page 24; 57pp; English.
XX
CC The present amino acid sequence is the protein-deamidating enzyme,
CC isolated from a new strain of soil bacterium, Chryseobacterium gleum
CC JCM 2410. This enzyme exerts the deamidation activity by directly acting
CC upon side chain amido groups in the protein in bonded state and
CC releasing side chain carboxyl groups and ammonia. It can deaminate high
CC molecular weight proteins, without cross linking and cleavage of peptide
CC bonds, to improve protein function. This sequence is used for protein
CC engineering, to cause an increase in surface hydrophobicity and improve
CC the function of a plant or animal protein. It can also be used to remove
CC or reduce toxicity or allergic property of proteins in food, decrease
CC mineral sensitivity of protein, to allow greater absorption into the body
CC and to solubilise calcium for use in drinks and mineral enhancing agents.
XX
SQ Sequence 185 AA;

Query Match 52.1%; Score 874; DB 21; Length 185;
Best Local Similarity 87.4%; Pred. No. 2.4e-74;
Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 138 SVIPLVATLNSLNFQIKNOSCTSTASSPCITFRFPVDCGYARAHKROIILMNGYDCCK 197
DB 3 sviplvatlnslnfqikngacgtstasspcitfrfpvdcgyarahkroiilmnagdyck 62
QY 198 OFPYGNLKAFTGTCVAMSYHVAIIIVSYKNASGVTEKRIIDPSLSSGPTDTAMRNA 257
DB 63 ofpygnlkaftgctcvawshvaailivsyknasgvtekrliidpslssgptdtawrna 122
QY 258 NTSKCSASVSYANTAGNYYRSPNSLYDNNLINTNCVLTFRFSLSGCSPSPADPVSS 317
DB 123 ntskcsasvsyantagnyyrspnslydnnlntncvltfrfslsgcspspadvss 182
QY 318 CGF 320
DB 183 cgf 185

```

RESULT      5
AA063793
ID   AAR63793 standard; Protein; 884 AA.
XX
AC   AAR63793;
XX
XX   12-JUL-1995 (first entry)
XX
DE   Bacillus cereus 100 kDa vegetative insecticidal protein-1.
XX
KM   Vegetative insecticidal protein-1; VIP-1; Bacillus cereus;
XX   pesticide; Diabrotica virgifera virgifera; insecticide.
XX
OS   Bacillus cereus.
XX
FT   Key                               Location/Qualifiers
FT   Misc-difference 217                /Label= Lys, Gln
FT   Misc-difference 245                /Label= Gln, His
XX
XX   MO9421795-A.
XX   PD   29-SEP-1994.
XX   XX   23-MAR-1994; 94WO-US03131.
XX   PF   25-MAR-1993; 93US-0037057.
XX   PR   (CIBA ) CIBA GEIGY AG.
XX   PA   Carr B, Desai N, Kostichka N, Koziel MG, Mullins MA;
XX   PI   Nye GJ, Warren GW;
XX   DR   WPI. 1994-317015/39.
XX   DR   N-PSDB: AA074679.
XX
PT   Novel pesticidal proteins and Bacillus strains - e.g. useful for
PT   control of Diabrotica virgifera virgifera
XX
PS   Claim 18; Page 63; 108pp; English.
XX
CC   AA074679 encodes AAR63793 the 100 kDa vegetative insecticidal protein-1
CC   (VIP-1), also claimed is the protein AAR63795 which enhances the
CC   pesticidal activity of VIP-1. The protein peptide combination can be
CC   used in a variety of systems for controlling plant and non-plant
CC   pests, including insects, fungi, bacteria, nematodes, protozoan
CC   pathogens and animal-parasitic liver flukes. However it is esp.
CC   useful in the control of Diabrotica virgifera virgifera.
XX
XX   Sequence 884 AA:
SQ
Query Match      6.9%; Score 116.5; DB 15; Length 884;
Best Local Similarity 23.5%; Pred. No. 0.088;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;
QY 16 TENSCADSNNGNOEIN-----GKEKLSYNDSKLDFGKTVPGVGDDENGMKIVSMFLTA 68
DB 114 tnlsedegallleingklisnkgkqvvhlek---gklvplkleygs---dtkfnlds 166
QY 69 Q-FYEIKPTKENEOYIGMLROAVKNESPVHIFLKPNSEIGVESASPEDVRYFTILTK 127
DB 167 ktfeklkikldsq--ngpqvgvqdelrnfefnkkesgef----lakpsklnlftqgmkr 220
QY 128 EVKGGOTNKLASVTPDVATLN--SLFNOIKNO-----SCG-TSTASSPCITFRYPVDCGY 178
DB 221 eldedtdtdgdsipdlweengytlxnrlavkwddslaskgycklfsnples--htvgdpy 278
QY 179 ARAHK-MROITLANNNGD-----CEKOFVGNLKAISGTGTCVAMSHVAIL 222

```

```

DB 279 tdyeakaardldlsnaketlnplvaafpsvsvsmekvllspnehlmsv-----eshstcn 333
QY 223 VSYKNASGVTEKRIIDPSLFSSG-----PVTDRAMRACVNTSCGSASVSANT-- 272
DB 334 wsyntegagsveagigpklstfsvsvnyqhselvaqegwstsgntsgntsfntasgylhany 393
QY 273 -----AGNYYRSPSPNSLYLDNNLINT 294
DB 394 rynnvgtgalydvkptcstfvlndtlat 421

RESULT      6
AAR91246
ID   AAR91246 standard; Protein; 852 AA.
XX
AC   AAR91246;
XX
XX   15-AUG-1996 (first entry)
XX
DE   VIP1A(a) protein with signal peptide removed.
XX
XX   Pesticide; insecticide; biological control agent; Lepidoptera;
XX   Coleoptera; transgenic plant; maize; Zea mays; insect resistance;
XX   VIP1; Bacillus cereus; western corn rootworm;
XX   Diabrotica vergifera vergifera.
XX
OS   Synthetic.
XX
XX   WO9610083-A1.
XX   PD   04-APR-1996.
XX   PF   27-SEP-1995; 95WO-EP03826.
XX   PR   05-JUN-1995; 95US-0463483.
XX   PR   28-SEP-1994; 94US-0314594.
XX   PA   (CIBA ) CIBA GEIGY AG.
XX   PI   Carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K;
XX   PI   Koziel MG, Mullinsma, Nye GJ, Warren GW;
XX   DR   WPI. 1996-200921/20.
XX   DR   N-PSDB: AAT13945.
XX
PT   Bacillus strain producing insecticidal protein during vegetative
PT   growth - used in the control of Lepidoptera and Coleoptera pests
XX
PS   Claim 31; Page 183-187; 242pp; English.
XX
CC   A version (AAR91246) of the Bacillus cereus strain AB78 VIP1A(a)
CC   insect-specific protein (see also AAR91239) lacks the Bacillus signal
CC   peptide and is the product of a synthetic gene (AAT13945) having codon
CC   usage optimised for expression in maize. VIP1A(a) shows activity
CC   against western corn rootworm. The synthetic gene can be fused to a
CC   synthetic gene (AAT13946) coding for the VIP2A(a) auxillary protein.
CC   The resulting construct (AAT13947) codes for a VIP2A(a)-VIP1A(a) fusion
CC   protein (AAR91247) that can be expressed in maize.
XX
XX   Sequence 852 AA:
SQ
Query Match      6.7%; Score 112.5; DB 17; Length 852;
Best Local Similarity 23.5%; Pred. No. 0.2; Mismatches 138; Indels 69; Gaps 15;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;
QY 16 TENSCADSNNGNOEIN-----GKEKLSYNDSKLDFGKTVPGVGDDENGMKIVSMFLTA 68
DB 82 tnlsedegallleingklisnkgkqvvhlek---gklvplkleygs---dtkfnlds 134
QY 69 Q-FYEIKPTKENEOYIGMLROAVKNESPVHIFLKPNSEIGVESASPEDVRYFTILTK 127
DB 135 ktfeklkikldsq--ngpqvgvqdelrnfefnkkesgef----lakpsklnlftqgmkr 188

```


CC against Sesamia pests. The method and compositions are especially used
CC for protecting maize but may also be used to protect other cereal crops
CC against Asian Corn Borer attack.

XX Sequence 852 AA:

Query Match 6.7%; Score 112.5; DB 19; Length 852;
Best Local Similarity 23.5%; Pred. No. 0.2;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

QY 16 TNSCADSNGNOEIN-----GKEKLSYNDKLRKFGKTVPGIDENGMIKVSFMLTA 68
DB 82 tnlsedegallieingklisnkyekqvvhlek----gklvplkieygs---dtkfnlds 134
QY 69 Q-FYEIKPTKENQYIGMLRQAVKNESPVHIFLKPSNNEIGKVESAPEDVRYFKTILRK 127
DB 135 ktfeklkrlkldsq--ngpqvqgdeldlnpefnkkesgef----lakpsklnlftqkmxr 188
QY 128 EVKGGTNRKLASYIPDVATLN--SLFNOIKNQ-----SCG-TSTASSPCITRRYPVDGCV 178
DB 189 eldedtdtdgdslpdlweengytlgnrlavkwddslaskytkfvsnples--htvgdpy 246
QY 179 ARAHK-MROIILNNGYD-----CEKQFVGNLKAStGTCVAMSYVALI 222
DB 247 tdyeKaardldsnaketlnplvaafpsvnmekvllspnenlnsv-----eshssten 301
QY 223 VSYKNASGVTEKRIIDPSLFSSG-----PYTDAMRNACVMTSCGSASVSYANT-- 272
DB 302 wylntegsveaglgpklisfgsvnyghsetvagekwtsfgntsgfntasagylmanv 361
QY 273 -----AGNVYRRSPNSLYLDNNLINT 294
DB 362 rynnvgtgalydvkprtsfvlnmdtiat 389

RESULT 9

AAM60224
ID AAM60224 standard; Protein: 880 AA.

XX AAM60224;

XX 28-SEP-1998 (first entry)

DE Bacillus thuringiensis insecticidal toxin 177C8.

KW Insecticide: pesticide: toxin; delta-endotoxin;

KM biological control; lepidopteran; coleopteran.

XX Bacillus thuringiensis strain PS177C8 (NRRL B-21867).

OS Key Location/Qualifiers

FT Misc-difference 253

FT Misc-difference 675 /note= "encoded by YMA"

FT Misc-difference 846 /note= "encoded by AC"

FT Misc-difference /note= "encoded by RAA"

XX W09818932-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19804.

XX 30-OCT-1996; 96US-0029848.

XX (MYCO) MYCOGEN CORP.

XX Dullum CJ, Feltelson JS, Loewer D, Muller-Cohn J;

XX Narva KE, Schmelts JL, Schneft HE, Schwab G, Stamp L;

XX PI Stockhoff BA;

DR WPI: 1998-272226/24.

DR N-PSDB; AAV30307.

XX Bacillus thuringiensis isolates - used for producing pesticidal

PT toxins and nucleotide sequences for control of lepidopterans and

PT coleopterans

PS Claim 5; Page 81-84; 139pp; English.

CC This polypeptide comprises a novel soluble toxin of Bacillus
CC thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin
CC belongs to a novel family of B.t. toxins that have toxicity
CC against non-mammalian pests. Its amino acid sequence was deduced
CC from a novel DNA fragment (see AAV30307) obtained by PCR from
CC cellular genomic DNA of PS177C8. Disclosed and claimed are novel
CC B.t. isolates and toxins (see AAM60218-32) that have activity against
CC lepidopteran and/or coleopteran pests, isolated genes, probes
CC and primers (see AAV30288-321 and AAV39734-87) useful for production
CC of the toxins and for the identification and characterisation of
CC these toxins, and transformed hosts, particularly plant and
CC bacterial hosts. The invention provides 8 entirely new families of
CC toxins from B.t. isolates. The toxins have the additional ability
CC to form pores in cell membranes, and can be used to facilitate
CC entry of a second agent into a target cell.

XX Sequence 880 AA:

Query Match 6.7%; Score 112.5; DB 19; Length 880;
Best Local Similarity 23.5%; Pred. No. 0.21;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

QY 16 TNSCADSNGNOEIN-----GKEKLSYNDKLRKFGKTVPGIDENGMIKVSFMLTA 68
DB 111 tnlsedegallieingklisnkyekqvvhlek----gklvplkieygs---dtkfnlds 163
QY 69 Q-FYEIKPTKENQYIGMLRQAVKNESPVHIFLKPSNNEIGKVESAPEDVRYFKTILRK 127
DB 164 ktfeklkrlkldsq--ngpqvqgdeldlnpefnkkesgef----lakpsklnlftqkmxr 217
QY 128 EVKGGTNRKLASYIPDVATLN--SLFNOIKNQ-----SCG-TSTASSPCITRRYPVDGCV 178
DB 218 eldedtdtdgdslpdlweengytlgnrlavkwddslaskytkfvsnples--htvgdpy 275
QY 179 ARAHK-MROIILNNGYD-----CEKQFVGNLKAStGTCVAMSYVALI 222
DB 276 tdyeKaardldsnaketlnplvaafpsvnmekvllspnenlnsv-----eshssten 330
QY 223 VSYKNASGVTEKRIIDPSLFSSG-----PYTDAMRNACVMTSCGSASVSYANT-- 272
DB 331 wylntegsveaglgpklisfgsvnyghsetvagekwtsfgntsgfntasagylmanv 390
QY 273 -----AGNVYRRSPNSLYLDNNLINT 294
DB 391 rynnvgtgalydvkprtsfvlnmdtiat 418

RESULT 10

AAV59277
ID AAV59277 standard; Protein: 881 AA.

XX AAV59277;

XX 18-APR-2000 (first entry)

DE MIS toxin from B. thuringiensis strain PS177C8.

XX Bacillus thuringiensis; toxin; endotoxin; pesticide; plant pest;

KW lepidopterans; coleopterans.

OS Bacillus thuringiensis.

XX Key

XX Location/Qualifiers

FT Misc-difference 846 /note= "unknown"

XX

PN W09957282-A2.

XX

PD 11-NOV-1999.

XX

PF 06-MAY-1999; 99WO-US09997.

XX

PR 06-MAY-1998; 98US-0073898.

XX

PA (MYCO) MYCOGEN CORP.

XX

PI Feltelson JS, Schnepf HE, Narva KE, Stockhoff BA, Schmeits J, Loewer D, Dullum CJ, Muller-Cohn J, Stamp L, Morrill G, Finstad-Lee S;

PI WPI: 2000-096811/08.

DR N-PSDB; AA258786.

XX

PT New polynucleotides encoding pesticidally active proteins, useful for transforming plants for controlling pests

XX

PS Disclosure: Page 68-71; 104pp; English.

XX

CC The invention relates to novel *B. thuringiensis* isolates, and genes encoding pesticidal toxins which are toxic to non-mammalian pests. The genes are useful in the control of non-mammalian pests and especially plant pests (e.g. lepidopterans and/or coleopterans). The polynucleotides are useful for transforming plants for controlling plant pests; for designing primers and probes useful for the identification and characterization of genes which encode pesticidal toxins. The present sequence represents a *B.t.* toxin.

CC

XX

SO Sequence 881 AA;

Query Match 6.7%; Score 112.5; DE 21; Length 881; Best Local Similarity 23.5%; Pred. No. 0.21; Mismatches 138; Indels 69; Gaps 15; Matches 77; Conservative 44;

QY 16 TFNSCADSNGNOEIN-----GKEKLSVNDSKLDFGKTYPVGIDEENGMIKYSFMLTA 68

DB 111 tfnlsedeqaliefngkllsnkgykvvhlek---gklvpikleygs---dtkfnids 163

QY 69 Q-FYEIKPTKEEQYIGMLRQAVKNESPVHIFLKPSNEIGKVESASPEDVRYEFTILT 127

DB 164 ktfelkikfkidsq--ngpqvgqdelrnpelnkkesgef---lakpsklnlftqkmkr 217

QY 128 EVKGOTNKLASYIPDVATLN--SLFNQIKNO-----SCG-TSTASSPCITFRYPVDCY 178

DB 218 eidedtdcdgsipdiweengytlgnrlavkwddsaskgytkfvsnples--hcvgqpy 275

QY 179 ARAHK-MROIIMNNGYD-----CEKQFYGNLKAStGTCVAMSYHVAIL 222

DB 276 tdyeakaardldlsnaketfnplvaafpsvnmekvllspenlsnv-----eshstn 330

QY 223 VSKKASGVTEKRIIDPDLFSSG-----PYDTARNMACVNTSCGSASVSYSTANT-- 272

DB 331 wsyntlegasveaglgpkyglsfgyvsnvqhsetvaqewgstlgnstgntasagylinanv 390

QY 273 -----ACNVYYRSPNSYLYDNNLINT 294

DB 391 rymnvgtgailgyvkpctsfvlnndliat 418

RESULT 11

AA91239

XX AA91239 standard; Protein: 884 AA.

AC

XX AA91239;

XX

DT 14-AUG-1996 (first entry)

XX

DE B. cereus VIP1a(a) insect-specific protein.

XX

KW Pesticide; insecticide; biological control agent; Lepidoptera; Coleoptera; transgenic plant; maize; insect resistance;

KW western corn rootworm; Diabrotica virgifera virgifera; VIP.

OS

XX Bacillus cereus strain AB78 (NRRL B-21058).

XX

FT Key Location/Qualifiers

FT Peptide 1..33

FT /Label= sig_peptide

XX

PN W09610083-A1.

XX

PD 04-APR-1996.

XX

PF 27-SEP-1995; 95WO-EP03826.

XX

PR 05-JUN-1995; 95US-0463483.

PR 28-SEP-1994; 94US-0314594.

XX

PA (CIBA) CIBA GEIGY AG.

XX

PI Carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K; Koziel MG, Mullinsma, Nye GJ, Warren GW;

PI WPI: 1996-200921/20.

DR N-PSDB; AAT13940.

XX

PT Bacillus strain producing insecticidal protein during vegetative growth - used in the control of Lepidoptera and Coleoptera pests

XX

PS Claim 15: Page 121-124; 242pp; English.

XX

CC Insect-specific protein VIP1a(a) (AA91239) of *Bacillus cereus* AB78 shows activity against *Diabrotica* spp. pests such as the western corn rootworm. It is encoded by the VIP1a(a) gene (AAT13940)

CC isolated from a cosmid clone of AB78. VIP1a(a) can be expressed in e.g. bacterial hosts to provide biological control agents having increased activity or target range, or can be expressed in transgenic plants, esp. maize, to improve insect resistance. It is preferably expressed as a fusion protein (see also AA91245) with auxiliary protein VIP2a(a) (AA91238).

CC

XX

SO Sequence 884 AA;

Query Match 6.7%; Score 112.5; DE 17; Length 884; Best Local Similarity 23.5%; Pred. No. 0.21; Mismatches 138; Indels 69; Gaps 15; Matches 77; Conservative 44;

QY 16 TFNSCADSNGNOEIN-----GKEKLSVNDSKLDFGKTYPVGIDEENGMIKYSFMLTA 68

DB 114 tfnlsedeqaliefngkllsnkgykvvhlek---gklvpikleygs---dtkfnids 166

QY 69 Q-FYEIKPTKEEQYIGMLRQAVKNESPVHIFLKPSNEIGKVESASPEDVRYEFTILT 127

DB 167 ktfelkikfkidsq--ngpqvgqdelrnpelnkkesgef---lakpsklnlftqkmkr 220

QY 128 EVKGOTNKLASYIPDVATLN--SLFNQIKNO-----SCG-TSTASSPCITFRYPVDCY 178

DB 221 eidedtdcdgsipdiweengytlgnrlavkwddsaskgytkfvsnples--hcvgqpy 278

QY 179 ARAHK-MROIIMNNGYD-----CEKQFYGNLKAStGTCVAMSYHVAIL 222

DB 279 tdyeakaardldlsnaketfnplvaafpsvnmekvllspenlsnv-----eshstn 333

QY 223 VSKKASGVTEKRIIDPDLFSSG-----PYDTARNMACVNTSCGSASVSYSTANT-- 272

DB 334 wsyntlegasveaglgpkyglsfgyvsnvqhsetvaqewgstlgnstgntasagylinanv 393

QY 273 -----ACNVYYRSPNSYLYDNNLINT 294

Query Match	6.78;	Score 112.5;	DB 17;	Length 1338;
Best Local Similarity	23.58;	Pred. No. 0.39;		
Matches 77;	Conservative 44;	Mismatches 138;	Indels 69;	Gaps 15;

CC application to the plant (or seed or growing area). The protein is
CC linear progeny, against insects of the genus *Sesamia* by direct or indirect
CC

CC especially useful to protect maize plants against the Mediterranean
CC corn borer (*S. nonagroides*).

xx
SQ Sequence 1338 AA;

Query Match 6.7%; Score 112.5; DB 18; Length 1338;
Best Local Similarity 23.5%; Pred. No. 0.39;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

```
QY 16 TFNSCADSNGNOEIN-----GKEKLSVNDSKLKDFTVPVGIDENGMIKVSFMLTA 68
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 568 tlnsedegalltelnqklsnkqkexqvvhlek---gklvpikleygs---dtkfnlds 620
QY 69 Q-FYEIKPTKENEOYIGMLRQAVKNESPVHIFLKPNSEIGKVESASPEDVRYFFKTLTK 127
   : ||| | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 621 ktfkelkfkidsq--nqpqyvgdelrnpetfnkkesgef----lakpsklnlftqgmkr 674
QY 128 EVKGGOTNKLASVTPDVATLN--SLENQIKNQ-----SCG-TSTASSPCITFRYPVDGCV 178
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 675 eldedtdtdgdsipdlweengytlqnrilavkwddsiasqkyckfvenples--htvgdpy 732
QY 179 ARAHK-MRQILMNNGVD-----CEKQFYVGNLKAStGTCVAMSYHVALT 222
   | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 733 tdyeakaardldisnaketlnplvaafpsvsvsmekvilsnpenlnsv-----eshstln 787
QY 223 VSYKNASGVTEKRIIDPSLFSSG-----PVTDTAMRNACVNTSCGSASVSYANT-- 272
   ||| | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 788 wsytnlegasveaglgpkglstfgvsvnyqhssetvagewgststgntsqfntasaagylnanv 847
QY 273 -----AGNVYYRSPSNSLYLDNNLINT 294
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 848 rynnvgtgalygvkptstfvlnndtlat 875
```

Search completed: June 27, 2002, 21:47:24
Job time: 5512 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:48:14 ; Search time 38.03 Seconds
(without alignments)
205.527 Million cell updates/sec

Title: US-09-727-769a-8
Perfect score: 1679
Sequence: 1 MKNLFSLMMAFVTLTFNSC.....FSLSGCSPSPADVSSCGF 320

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1292.5	77.0	319	US-09-324-910-12	Sequence 12, Appl
2	874	52.1	185	US-09-324-910-6	Sequence 6, Appl
3	112.5	6.7	852	US-08-471-033-36	Sequence 36, Appl
4	112.5	6.7	852	US-08-471-044-36	Sequence 36, Appl
5	112.5	6.7	852	US-08-463-483A-36	Sequence 36, Appl
6	112.5	6.7	852	US-08-471-046A-36	Sequence 36, Appl
7	112.5	6.7	852	US-08-470-366B-36	Sequence 36, Appl
8	112.5	6.7	852	US-08-469-334-36	Sequence 36, Appl
9	112.5	6.7	852	US-09-300-529-36	Sequence 36, Appl
10	112.5	6.7	881	US-08-960-780-32	Sequence 32, Appl
11	112.5	6.7	881	US-09-073-898-32	Sequence 32, Appl
12	112.5	6.7	884	US-08-471-033-5	Sequence 5, Appl
13	112.5	6.7	884	US-08-471-044-5	Sequence 5, Appl
14	112.5	6.7	884	US-08-463-483A-5	Sequence 5, Appl
15	112.5	6.7	884	US-08-471-046A-5	Sequence 5, Appl
16	112.5	6.7	884	US-08-470-566B-5	Sequence 5, Appl
17	112.5	6.7	884	US-08-469-334-5	Sequence 5, Appl
18	112.5	6.7	884	US-09-300-529-5	Sequence 5, Appl
19	112.5	6.7	1338	US-08-471-033-50	Sequence 50, Appl
20	112.5	6.7	1338	US-08-471-044-50	Sequence 50, Appl
21	112.5	6.7	1338	US-08-463-483A-50	Sequence 50, Appl
22	112.5	6.7	1338	US-08-471-046A-50	Sequence 50, Appl
23	112.5	6.7	1338	US-08-470-566B-50	Sequence 50, Appl
24	112.5	6.7	1338	US-08-469-334-50	Sequence 50, Appl
25	112.5	6.7	1338	US-09-300-529-50	Sequence 50, Appl
26	112.5	6.7	1346	US-08-471-033-23	Sequence 23, Appl
27	112.5	6.7	1346	US-08-471-044-23	Sequence 23, Appl

28	112.5	6.7	1346	US-08-463-483A-23	Sequence 23, Appl
29	112.5	6.7	1346	US-08-471-046A-23	Sequence 23, Appl
30	112.5	6.7	1346	US-08-470-566B-23	Sequence 23, Appl
31	112.5	6.7	1346	US-08-469-334-23	Sequence 23, Appl
32	112.5	6.7	1346	US-09-300-529-23	Sequence 23, Appl
33	95	5.7	1541	US-08-296-791-3	Sequence 3, Appl
34	95	5.7	1541	PCT-US95-10661A-3	Sequence 3, Appl
35	93	5.5	1183	US-08-447-031A-2	Sequence 2, Appl
36	91.5	5.4	615	US-08-525-742-10	Sequence 10, Appl
37	90.5	5.4	1222	US-08-682-517-15	Sequence 15, Appl
38	90.5	5.4	1252	US-08-682-517-9	Sequence 9, Appl
39	89.5	5.3	610	US-08-525-742-8	Sequence 8, Appl
40	89.5	5.3	1338	US-08-728-470-9	Sequence 9, Appl
41	89.5	5.3	1338	US-08-719-641-9	Sequence 9, Appl
42	89.5	5.3	1599	US-08-617-697-9	Sequence 9, Appl
43	88.5	5.3	512	US-08-856-253-6	Sequence 6, Appl
44	88	5.2	834	US-08-471-033-21	Sequence 21, Appl
45	88	5.2	834	US-08-471-044-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1	US-09-324-910-12	Sequence 12, Application US/09324910
Patent No. 6251651		
GENERAL INFORMATION:		
APPLICANT: Yamaguchi, Shotaro		
APPLICANT: Matsura, Akira		
TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE		
TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF		
FILE REFERENCE: 0-54362		
CURRENT APPLICATION NUMBER: US/09/324,910		
CURRENT FILING DATE: 1999-06-03		
EARLIER APPLICATION NUMBER: HEI-10-173940		
EARLIER FILING DATE: 1998-06-04		
NUMBER OF SEQ ID NOS: 12		
SOFTWARE: PatentIn Ver. 2.0		
SEQ ID NO 12		
LENGTH: 319		
TYPE: PRT		
ORGANISM: Chryseobacterium gleum		
US-09-324-910-12		
Query Match	77.0%, Score 1292.5, DB 4, Length 319;	
Best Local Similarity	74.7%, Pred. No. 5.5e-129;	
Matches 239; Conservative 39; Mismatches 41; Indels 1; Gaps 1;		
QY 1 MKNLFSLMMAFVTLTFNSCADSNGNOELNGEKLSVNDSKRKDGKTVPGIDENGMI 60		
1 MKNLFSLMMAFVTLTFNSCADSNGNOELNGEKLSVNDSKRKDGKTVPGIDENGMI 59		
QY 61 KVSFELTFOFEYIKETKTEQNYIGMLROAVKNSEPVHIFLKPNSNEIKVESASPEDEVY 120		
61 KVSFELTFOFEYIKETKTEQNYIGMLROAVKNSEPVHIFLKPNSNEIKVESASPEDEVY 119		
QY 60 KISFVNTQAPPEIADSKENAGTSMIRQAVENETVHVHFLKNTKIKAKVEATDDDIRY 119		
60 KISFVNTQAPPEIADSKENAGTSMIRQAVENETVHVHFLKNTKIKAKVEATDDDIRY 119		
QY 121 FKTITTKVKQGTNKLASVIVPVATLNSLFNOIKQSGCTSTASSPCITFRYPVDCYAR 180		
121 FKTITTKVKQGTNKLASVIVPVATLNSLFNOIKQSGCTSTASSPCITFRYPVDCYAR 179		
QY 181 AHKMQILMNNNGYDEKQFVYGNLKSASGTCCVAAHYVALIVSKNAGVTEKRTIDPS 240		
181 AHKMQILMNNNGYDEKQFVYGNLKSASGTCCVAAHYVALIVSKNAGVTEKRTIDPS 239		
QY 241 LFSGSPVDTAMRNACVMTSCGASASVSYANTAGVYRSPNSLYLNNTINTNCVLTG 300		
241 LFSGSPVDTAMRNACVMTSCGASASVSYANTAGVYRSPNSLYLNNTINTNCVLTG 299		
QY 301 FSLSGCSPSPADVSSCGF 320		
301 FSLSGCSPSPADVSSCGF 319		
Db		

RESULT 2
US-09-324-910-6
Sequence 6, Application US/09324910
Patent No. 6251651
GENERAL INFORMATION:
APPLICANT: Yamaguchi, Shotaro
APPLICANT: Matsura, Akira
TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
FILE REFERENCE: O-54362
CURRENT APPLICATION NUMBER: US/09/324, 910
CURRENT FILING DATE: 1999-06-03
EARLIER APPLICATION NUMBER: HEI-10-173940
EARLIER FILING DATE: 1998-06-04
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 185
TYPE: PRT
ORGANISM: Chryseobacterium gleum
US-09-324-910-6

Query Match 52.1%; Score 874; DB 4; Length 185;
Best Local Similarity 87.4%; Pred. No. 6.8e-85;
Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 138 SVIPDVATLNSLFENQKNSCGTSTASPCITFRYPVDCYARAHKROILMNGYDCER 197
DB 3 SVIPDLATLNSLFQTKNOACGTSTASPCITFRYPVDCYARAHKROILMAGTDCER 62
QY 198 QFVYGNLKAATGTCVAVMSYHVALVSYKNASGVTEKRIIDPSLFSSGCVDTAMNACV 257
DB 63 QFVYGNLKAATGTCVAVMSYHVALVSYKNASGVTEKRIIDPSLFSSGCVDTAMNACV 122
QY 258 MNSCGSASVSTANRAGNVYRSPNSLYLDNNLINTNCVLRKFLSLSCGSPAPDVSS 317
DB 123 MNSCGSASVSTANRAGNVYRSPNSLYLDNNLINTNCVLRKFLSLSCGSPAPDVSS 182
QY 318 CGF 320
DB 183 CGF 185

RESULT 3
US-08-471-033-36
Sequence 36, Application US/08471033
Patent No. 5770696
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 852 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-033-36

Query Match 6.7%; Score 112.5; DB 1; Length 852;
Best Local Similarity 23.5%; Pred. No. 0.0076;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

QY 16 TFSQADNSNGNOEIN-----GKEKLSYNDKLFQKTPVQIDEGEMKIVSFMLTA 68
DB 82 TNLSEDEQAIITFINKITSNKKEQVYHLEK---GLVPIKIEYQS---DTKFNIDS 134
QY 69 Q-FYEIKPTKENEQYIGLMROAVKNESPVHIFLKPSNEIGKVESASPEDVRYFKTILRK 127
DB 135 KTFKELKFKIDSQ--NQOQVOQODELRNPFKKKESGEF---LAKPSKINLFTQOMKR 188
QY 128 EVKQGTNKLASYIPDVATLN--SLFNOIKNQ-----SCG-TSTASPCITFRYPVDCY 178
DB 189 EIDEDPTDGDSPDLMEENGTYIQNRIVAKWDDSLASKGYTFVSNPLES--HTVGDPY 246
QY 179 ARAHK-MROILMNGYD-----CEKQFVYGNLKAATGTCVAVMSYHVALI 222
DB 247 TDYERKARDLDSNAKETNPPLVAARFPVSVNSMEKYIILSPNERLSNV-----ESHSTN 301
QY 223 VSYKNASGVTEKRIIDPSLFSSG-----PYTDAMRNACVNTSCGSASVSYANT-- 272
DB 302 WSYTNEGASVEAGIGPKISFGVSVNYQHSETVAGDEMGTSTGNTSQFTASAGYINANV 361
QY 273 -----AGNVYRSPNSLYLDNNLINT 294
DB 362 RYNNVGTGAIYDVKKPTTSFLNNDTAT 389

RESULT 4
US-08-471-044-36
Sequence 36, Application US/08471044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J

Db 189 E1DEDTDGDS1PDLMEENGYTIONRIAYKMDDSLASKGYTKFVSNPLES--HTVGDPI 246
QY 179 ARAHK-MROIILMNGYD-----CEKOFVYGNLKAStGTCCVAMSYHVALI 222
| | : : : | | : : : |
Db 247 TDEKARLDLSNMKEFTNPVLAAPSVNVMKEKYLSPNEMLNSV-----ESHSSIN 301
QY 223 VSKNASGYTEKRIIDPSLFSSG-----PYDTAMRNACVNTSCGSASVSSTANT-- 272
| | : : : | | : : : |
Db 302 WSYTNEGASVAGIGPKGISFGVSVNYOHSETVAOEWGCTSGNTSOFMTASAGYLNANV 361
QY 273 -----AGNVYRRSPNSYLYDNNTLINT 294
| | : : : | | : : : |
Db 362 RYNNVGTGAIYDKPPTTSFVLNNDTIAT 389

RESULT 6
US-08-471-046A-36
; Sequence 36, Application US/08471046A
; Patent No. 5866326
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Method for Isolating Vegetative Insecticidal
; TITLE OF INVENTION: Protein Genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866326artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,046A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 852 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-471-046A-36
Query Match 6.7%; Score 112.5; DB 2; Length 852;
Best Local Similarity 23.5%; Pred No. 0.0076;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;
QY 16 TNSCADSNGNOEIN-----GKEKLSVNSDKLDPFGTVPVGDIEENGMIKVSFMLTA 68
| | : : : | | : : : |
Db 82 TNLSEDEQAIIENKGIISNKGKEQVHLEK---GLVPIKIEIKYQ---DTKENIDS 134
QY 69 Q-FYEIKPTKEEDQYIGMLRQAVKNESPVHIFLKPSNIEIGKVESAPEDVRYFKITLTK 127
| | : : : | | : : : |
Db 135 KTFEKLKLFKIDSQ--NOPQVOODELRNPFENKKEQOE---LAKPKRINLFTQOMKR 188
QY 128 EVKGFNTKLASVLPVATLN--SLFMQIKNO-----SCG-TSTASSPCTFRYPVGDGY 178
| | : : : | | : : : |
Db 189 E1DEDTDGDS1PDLMEENGYTIONRIAYKMDDSLASKGYTKFVSNPLES--HTVGDPI 246
QY 179 ARAHK-MROIILMNGYD-----CEKOFVYGNLKAStGTCCVAMSYHVALI 222
| | : : : | | : : : |
Db 247 TDEKARLDLSNMKEFTNPVLAAPSVNVMKEKYLSPNEMLNSV-----ESHSSIN 301
QY 223 VSKNASGYTEKRIIDPSLFSSG-----PYDTAMRNACVNTSCGSASVSSTANT-- 272
| | : : : | | : : : |
Db 302 WSYTNEGASVAGIGPKGISFGVSVNYOHSETVAOEWGCTSGNTSOFMTASAGYLNANV 361
QY 273 -----AGNVYRRSPNSYLYDNNTLINT 294
| | : : : | | : : : |
Db 362 RYNNVGTGAIYDKPPTTSFVLNNDTIAT 389

RESULT 7
US-08-470-566B-36
; Sequence 36, Application US/08470566B
; Patent No. 5872212
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5872212el pestiocidal Proteins and Strains
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5872212artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,566B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 852 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-566B-36

Query Match 6.7%; Score 112.5; DE 2; Length 852;
Best Local Similarity 23.5%; Pred. No. 0.0076;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

QY 16 TFNSCADSNGNOEIN-----GKEKLSVND SKLKD FGYVPGIDEENGMIKYSFMLTA 68
Db 82 TFNLSEDOAIIEINGKIISNKGKEKYVHLEK---GKLVPIKIEYQS---DFKFNIDS 134

QY 69 Q-FYEIRKPTKENEQYIGMLROAVKNESPVHIFLKPNSEIGKVESASPEDVRYEFTILT 127
Db 135 KTFEKLKLFKIDSO--NPOQVOODELRNPFENKKESEF---LAKSKINLFTQOKMR 188

QY 128 EVKQTNKLASVDPVATLN--SLFNQIKNO-----SCG-TSTASSPCITFRYPVDCY 178
Db 189 EIDEDTDTGDSIPDLMEENGTYIONRIAVKWDLSLASKGYTKFVSNPLES--HTVGDPY 246

QY 179 ARAHK-MQOILMNNGYD-----CEKQFYGNLKA STGTCVAMS YHVAIL 222
Db 247 TDYEKAARDLDSNAKEFNPLVAAFP SVNVM EKVVILSPENLNSV-----ESHSTN 301

QY 223 VSKNASGVTEKRIIDPLFSFG-----PVDTARNACVNTSCGSASVSYANT-- 272
Db 302 WSYNTBEGASVEAGIGPGISFGVSVNYOHSETVAQEWGISTGNTSOPNTASAGYLANV 361

QY 273 -----AGNYYRSPSN SYLYDNMLINT 294
Db 362 RYNNVGTGAIVDKFTTSFVLNNDTIAT 389

RESULT 8
US-08-469-334-36
Sequence 36, Application US/08469334
Patent No. 5990383
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,334
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/463,483
FILING DATE: US 08/314,594
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 852 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-334-36

Query Match 6.7%; Score 112.5; DB 2; Length 852;
Best Local Similarity 23.5%; Pred. No. 0.0076;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

QY 16 TFNSCADSNGNOEIN-----GKEKLSVND SKLKD FGYVPGIDEENGMIKYSFMLTA 68
Db 82 TFNLSEDOAIIEINGKIISNKGKEKYVHLEK---GKLVPIKIEYQS---DFKFNIDS 134

QY 69 Q-FYEIRKPTKENEQYIGMLROAVKNESPVHIFLKPNSEIGKVESASPEDVRYEFTILT 127
Db 135 KTFEKLKLFKIDSO--NPOQVOODELRNPFENKKESEF---LAKSKINLFTQOKMR 188

QY 128 EVKQTNKLASVDPVATLN--SLFNQIKNO-----SCG-TSTASSPCITFRYPVDCY 178
Db 189 EIDEDTDTGDSIPDLMEENGTYIONRIAVKWDLSLASKGYTKFVSNPLES--HTVGDPY 246

QY 179 ARAHK-MQOILMNNGYD-----CEKQFYGNLKA STGTCVAMS YHVAIL 222
Db 247 TDYEKAARDLDSNAKEFNPLVAAFP SVNVM EKVVILSPENLNSV-----ESHSTN 301

QY 223 VSKNASGVTEKRIIDPLFSFG-----PVDTARNACVNTSCGSASVSYANT-- 272
Db 302 WSYNTBEGASVEAGIGPGISFGVSVNYOHSETVAQEWGISTGNTSOPNTASAGYLANV 361

QY 273 -----AGNYYRSPSN SYLYDNMLINT 294
Db 362 RYNNVGTGAIVDKFTTSFVLNNDTIAT 389

RESULT 9
US-09-300-529-36
Sequence 36, Application US/09300529
Patent No. 6066783
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian

Db 164 KTFEELKLFKIDSQ--NOPOOVODELRNPEFNKKESEOF-----LAKPSKINLFTQMKR 217
 QY 128 EVKQGTNKLASVIPPVATLN--SLFNQIKNQ-----SCG-TSTASSPCITFRYPVDCY 178
 Db 218 EIDEDTDTGDSIPDLMEENGCTTQNRIVAKWDSLASKGYTKFVSNPLES--HTVGDPY 275
 QY 179 ARAHK-MRQILMNNGYD-----CEKQFYVGNLKAStGTCVAMSYHVAIL 222
 Db 276 TDYEKAARDLDSNAKEFTFNPLVAFPSVNVSMKEVILSPENLNSV-----ESHSTN 330
 QY 223 VSYKNASGVTEKRIIDPSLFSSG-----PYDTAMRNACVNTSCGSASVSYANT-- 272
 Db 331 WSYTTEGASVEAGIGPKGISFGVSVNYQHSETVAOEWGTSGTNTSOPNTASAGYLANV 390
 QY 273 -----AGNVYRSPNSLYDNMLINT 294
 Db 391 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 418

RESULT 11
 US-09-073-898-32
 : Sequence 32, Application US/09073898
 : Patent No. 6242669
 : GENERAL INFORMATION:
 : APPLICANT: Feltelson, Jerald S.
 : APPLICANT: Schepf, H. Ernest
 : APPLICANT: Narva, Kenneth E.
 : APPLICANT: Stochhoff, Brian A.
 : APPLICANT: Schmelts, James
 : APPLICANT: Loewer, David
 : APPLICANT: Dullum, Charles Joseph
 : APPLICANT: Muller-Cohn, Judy
 : APPLICANT: Stamp, Lisa
 : APPLICANT: Mortill, George
 : APPLICANT: Finstad-Lee, Stacey
 : TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
 : TITLE OF INVENTION: Sequences Which Encode These Toxins
 : NUMBER OF SEQUENCES: 144
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
 : STREET: 2421 N.W. 41st Street, Suite A-1
 : CITY: Gainesville
 : STATE: FL
 : COUNTRY: US
 : ZIP: 32606-6669
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/073.898
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 60/029,848
 : FILING DATE: 30-OCT-1996
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/960,780
 : FILING DATE: 30-OCT-1997
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Sanders, Jay M.
 : REGISTRATION NUMBER: 39,355
 : REFERENCE/DOCKET NUMBER: MA-708C1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 352-375-8100
 : TELEFAX: 352-372-5800
 : INFORMATION FOR SEQ ID NO: 32:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 881 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: peptide
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: PS177C8
 US-09-073-898-32

Query Match 6.7%; Score 112.5; DB 4; Length 881;
 Best Local Similarity 23.5%; Pred. No. 0.008;
 Matches 77; Conservative 44; Mismatches 18; Indels 69; Gaps 15;

QY 16 TFNSCADSNGNOETN-----GKEKLSVNDSKLDEKGTVPVGDIDENGMIKVSFMLTA 68
 Db 111 TFNLSEDOALIEITNGKIISNKGKEKQVHLEK-----GKLVPKIRIEYOS--DTRFNIDS 163
 QY 69 Q-FYEIKRTKNEQYIGMLRQAVKNESVHIFELPNSNEIKKVSASSEDYRKYTKILT 127
 Db 164 KTFEELKLFKIDSQ--NOPOOVODELRNPEFNKKESEOF-----LAKPSKINLFTQMKR 217
 QY 128 EVKQGTNKLASVIPPVATLN--SLFNQIKNQ-----SCG-TSTASSPCITFRYPVDCY 178
 Db 218 EIDEDTDTGDSIPDLMEENGCTTQNRIVAKWDSLASKGYTKFVSNPLES--HTVGDPY 275
 QY 179 ARAHK-MRQILMNNGYD-----CEKQFYVGNLKAStGTCVAMSYHVAIL 222
 Db 276 TDYEKAARDLDSNAKEFTFNPLVAFPSVNVSMKEVILSPENLNSV-----ESHSTN 330
 QY 223 VSYKNASGVTEKRIIDPSLFSSG-----PYDTAMRNACVNTSCGSASVSYANT-- 272
 Db 331 WSYTTEGASVEAGIGPKGISFGVSVNYQHSETVAOEWGTSGTNTSOPNTASAGYLANV 390
 QY 273 -----AGNVYRSPNSLYDNMLINT 294
 Db 391 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 418

RESULT 12
 US-08-471-033-5
 : Sequence 5, Application US/08471033
 : Patent No. 5770696
 : GENERAL INFORMATION:
 : APPLICANT: Warren, Gregory W
 : APPLICANT: Kozziel, Michael G
 : APPLICANT: Mullins, Martha A
 : APPLICANT: Nye, Gordon J
 : APPLICANT: Carr, Brian
 : APPLICANT: Desai, Nalini M
 : APPLICANT: Kostichka, N. Kristy
 : APPLICANT: Duck, Nicholas B
 : APPLICANT: Estruch, Juan J
 : TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
 : NUMBER OF SEQUENCES: 50
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: CIBA-GEIGY Corporation
 : STREET: 7 Skyline Drive
 : CITY: Hawthorne
 : STATE: NY
 : COUNTRY: USA
 : ZIP: 10532
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30B
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/471.033
 : FILING DATE:
 : CLASSIFICATION: 530
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/314,594
 : FILING DATE: 09-SEP-1994
 : APPLICATION DATA:
 : APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-033-5

Query Match 6.7%; Score 112.5; DB 1; Length 884;
Best Local Similarity 23.5%; Pred. No. 0.0081;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

QY 16 TNSCADSNGNDEIN-----GKELSYNDSKLDGKTVPGVIGDEENGMKIVSFMULTA 68
DB 114 TNLSEDEQALIEINKKISNKGKQVYHLER---GLVPIKIEYQS---DTKFNIDS 166
QY 69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPDVRYFKTILRK 127
DB 167 KTFKELKLFKIDSQ--NQOQVOQODELRNPFENKKSQEF---LAKPSKINLFTQKKMR 220
QY 128 EVKGGTNNKLASYIPDVATIN--SLENOIKNQ-----SCG-TSTASSPCITFRYPVDCGY 178
DB 221 EIDEPTDITDGDSDIPDLMEENGYTIONRIAVKMDDSLASKGYTKFVSNPLES--HTVGDpy 278
QY 179 ARAHK-MRQILNNGCYD-----CEKQFYGNLKAISGTCGVAMSHVAIL 222
DB 279 TDYEKAARDLDSNAKETFPNPLVAAPPSVNVSMKEVILSPNNLSNV-----ESHSSYN 333
QY 223 VSKNASGVTEKRIIDPSLFSSG-----PVDTAMRNACVNTSCGSASVSYANT-- 272
DB 334 WSYTTEGASVAGIGPKISFGVSVNYQHSETVAGQEMGTSTGNTSQFMTASAGYLNANV 393
QY 273 -----AGNVYRSPSNSLYLDNNLINT 294
DB 394 RYNNVGTGAIYDVKKPTTSFVLNNDTIAT 421

RESULT 13
US-08-471-044-5
Sequence 5, Application US/08471044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory M
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estrich, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-044-5

Query Match 6.7%; Score 112.5; DB 2; Length 884;
Best Local Similarity 23.5%; Pred. No. 0.0081;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;

QY 16 TNSCADSNGNDEIN-----GKELSYNDSKLDGKTVPGVIGDEENGMKIVSFMULTA 68
DB 114 TNLSEDEQALIEINKKISNKGKQVYHLER---GLVPIKIEYQS---DTKFNIDS 166
QY 69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIFLKPNSNEIGKVESASPDVRYFKTILRK 127
DB 167 KTFKELKLFKIDSQ--NQOQVOQODELRNPFENKKSQEF---LAKPSKINLFTQKKMR 220
QY 128 EVKGGTNNKLASYIPDVATIN--SLENOIKNQ-----SCG-TSTASSPCITFRYPVDCGY 178
DB 221 EIDEPTDITDGDSDIPDLMEENGYTIONRIAVKMDDSLASKGYTKFVSNPLES--HTVGDpy 278
QY 179 ARAHK-MRQILNNGCYD-----CEKQFYGNLKAISGTCGVAMSHVAIL 222
DB 279 TDYEKAARDLDSNAKETFPNPLVAAPPSVNVSMKEVILSPNNLSNV-----ESHSSYN 333
QY 223 VSKNASGVTEKRIIDPSLFSSG-----PVDTAMRNACVNTSCGSASVSYANT-- 272
DB 334 WSYTTEGASVAGIGPKISFGVSVNYQHSETVAGQEMGTSTGNTSQFMTASAGYLNANV 393
QY 273 -----AGNVYRSPSNSLYLDNNLINT 294
DB 394 RYNNVGTGAIYDVKKPTTSFVLNNDTIAT 421

RESULT 14
US-08-463-483A-5
Sequence 5, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory M
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE: 09-SEP-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-483A-5

Query Match 6.7%; Score 112.5; DB 2; Length 884;
Best Local Similarity 23.5%; Pred. No. 0.0081;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;
QY 16 TENSACADNGNOEIN-----GKEKLSVNSDKLDFGKTVPGVIGDENGMKIVSFMJTA 68
Db 114 TENISEDOALIEINGKIISNKGKEQYVHLEK---GKLVPIKIEYOS---DTKFNIDS 166
QY 69 Q-FYEIRPTKNEQYIGLRQAVKNESVHIFLKPNSMETKVSASASEDVRYRTITLTK 127
Db 167 KTFEKLFLFKIDSQ--NOPOQVODELNNPFRNKESEOF---LAKSKINLFTQKKR 220
QY 128 EVKQGTNKLASVDPVATLN--SLFNOIKNO-----SCG--TSTAASSPCIFRYPVDCY 178
Db 221 EIDEDTIDGSIPLDMEENGTITONRIAYKWDLSLASKGTYKVSNPLES--HTVGDPY 278
QY 179 ARAHK-MRQILMNNGYD-----CEKQFYGNLKAJSTGTCVAVSYHAIL 222
Db 279 TDYKKAARDLDSNAKETFNPLVAFPSVNVSMKVKVILSPNENLSNV-----ESHSTN 333
QY 223 VSYKNSGVTKRRIIDPSLFSSG-----PVDTAMRNACVNTSCGSASVSYANT-- 272
Db 334 WSYNTGASVSEAGIGPGISFVSVNYOHSETVAQEWGTSTGNTSOPNTASAGYLANV 393

QY 273 -----AGNYYRSPNSNYLVNMLINT 294
Db 394 RYNNVGTGAIDYKVPKPTTSFVLNNDTIAT 421

RESULT 15
US-08-471-046A-5
Sequence 5, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
TITLE OF INVENTION: Protein Genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5866326artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLv4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-046A-5

Query Match 6.7%; Score 112.5; DB 2; Length 884;
Best Local Similarity 23.5%; Pred. No. 0.0081;
Matches 77; Conservative 44; Mismatches 138; Indels 69; Gaps 15;
QY 16 TENSACADNGNOEIN-----GKEKLSVNSDKLDFGKTVPGVIGDENGMKIVSFMJTA 68
Db 114 TENISEDOALIEINGKIISNKGKEQYVHLEK---GKLVPIKIEYOS---DTKFNIDS 166

```

OY 69 Q-FYEIKPTKENEQYIGMLRQAVKNESPVHIEFLKPNSEIGKVESASPEDVRYFKTILTK 127
Db 167 KTFKELKLFKIDSQ--NOPOQVODELRNPFENKESQEF---LAKPSKINLFTOKMKR 220
OY 128 EYKGOTNKIASYIPVATIN--SLENQIKNO-----SCG-TSTASSPCITFRYPVDCY 178
Db 221 EIDEDTDIDGDSIPDLWEENGYTIONRIAWKDDSLASKGYTKFVSNPLES--HTVGDY 278
OY 179 ARAHK-MROIANNNGYD-----CEKOFVYGNLKAJSTGTCCVAMSYHAIL 222
Db 279 TDYEKAARDLDSNAKFTFNPLVAAFPSPVNSMEKYIILSPNENLSNV-----ESHSTN 333
OY 223 VSYKNASGVTEKRIIDPSLFSSG-----PVDTAMRNACVNTSCGSASVSYANT-- 272
Db 334 WSYTWTGASVEAGIGPKGISFPGSVNYQHSFVAQEMGTSTGNTSQFNTASAGYLNANV 393
OY 273 -----AGNVYYSRSPSNSLYLDNNLINT 294
Db 394 RYNNVGTGAIYDVKPTTSFVLNNDTIAT 421

```

Search completed: June 27, 2002, 21:48:16
 Job time: 5364 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:49:25 ; Search time 55.76 Seconds
(Without alignments)
551.445 Million cell updates/sec

Title: US-09-727-769a-8
Perfect score: 1679
Sequence: 1 MNKFLSMWAFVTILTFNSC.....FSLSGSCSPAPDVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	6.4	927	2 T21772	hypothetical prote
2	101.5	6.0	551	2 A83232	conserved hypothet
3	101	6.0	2395	1 S50820	surface protein ty
4	99.5	5.9	517	1 G1BP74	gene 12 protein -
5	99.5	5.9	534	2 T39903	serine-rich protei
6	99	5.9	298	2 I64138	adhesin homolog HI
7	99	5.9	564	2 J01792	Salp17R protein -
8	99	5.9	594	2 T50765	adhesion of calyx
9	98.5	5.9	419	2 C83912	hypothetical prote
10	97	5.8	534	2 C69146	hypothetical prote
11	96.5	5.7	260	2 B90026	hypothetical prote
12	96.5	5.7	490	2 H70103	hypothetical prote
13	95.5	5.7	1679	2 C42523	probable membrane
14	95	5.7	564	2 C42523	probable membrane
15	95	5.7	1541	2 A37023	15S protein - vac
16	94.5	5.6	622	2 A64494	15S-specific metal
17	93.5	5.6	1894	2 JC4980	hypothetical prote
18	93	5.5	594	2 T50764	plexin 1 precursor
19	92.5	5.5	1582	2 AC1153	adhesion of calyx
20	92	5.5	1154	2 A86318	adhesin homolog 1m
21	91.5	5.4	627	2 B64710	protein P15H18.11
22	91.5	5.4	851	2 S67285	site-specific DNA-
23	91.5	5.4	2013	2 A11489	ND1 protein - yea
24	91	5.4	396	2 T52596	probable peptidogl
25	91	5.4	761	1 TJHUNG	squamosa promoter
26	90.5	5.4	843	2 T22182	neural cell adhesi
27	90	5.4	243	2 T28802	hypothetical prote
28	89.5	5.3	1578	2 AD1512	hypothetical prote
29	89.5	5.3	2500	2 G71609	peptidoglycan boun
					hypothetical prote

30	89.5	5.3	2510	2 T28160	hypothetical prote
31	89	5.3	595	2 B86798	prophage p13 prote
32	89	5.3	1073	2 D87596	tonB-dependent rec
33	89	5.3	1895	2 T15881	hypothetical prote
34	88.5	5.3	283	2 G69212	conserved hypothet
35	88.5	5.3	466	2 H96991	secreted protein c
36	88.5	5.3	491	2 T20061	hypothetical prote
37	88.5	5.3	596	2 AC1318	two-component sens
38	88.5	5.3	685	2 AF0850	hypothetical prote
39	88.5	5.3	693	2 T15728	hypothetical prote
40	88.5	5.3	1251	2 T21389	hypothetical prote
41	88.5	5.3	1341	2 S66835	probable membrane
42	88	5.2	603	2 AE2119	GTP-binding elonga
43	88	5.2	814	2 T02195	hypothetical prote
44	88	5.2	2013	2 AD1129	probable peptidogl
45	87.5	5.2	372	2 F86189	hypothetical prote

ALIGNMENTS

RESULT 1
T21772
hypothetical protein T01D3.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21772, T24296
R:Steward, C.
Submitted to the EMBL Data Library, November 1996
A:Reference number: Z19470
A:Accession: T21772
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-927 <W12>
A:Cross-references: EMBL:Z81527; PIDN:CAB04279.1; GSPDB:GN00023; CESP:T01D3.6
A:Experimental source: clone F35E12
R:Steward, C.
Submitted to the EMBL Data Library, October 1996
A:Reference number: Z19870
A:Accession: T24296
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-927 <W12>
A:Cross-references: EMBL:Z81110; PIDN:CAB03262.1; GSPDB:GN00023; CESP:T01D3.6
A:Experimental source: clone T01D3
C:Genetics:
A:Gene: CESP:T01D3.6
A:Map position: 5
A:Introns: 38/2; 87/3; 128/2; 242/2; 374/3; 416/3; 734/3; 842/3; 857/1

Query Match 6.4%; Score 108; DB 2; Length 927;
Best Local Similarity 19.2%; Pred. No. 2.8;
Matches 85; Conservative 47; Mismatches 101; Indels 210; Gaps 24;

OY 35 LSVNDSKLDKFGKTVPV-----GIDENGMIKVSFMLAOFYEIKPTRENGOYI 83
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 15 LAABEDLKCYGADCEVHYMDTBOGSGLECSE--KETILAMISHYQ---KEKHDTYL 68
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 84 GMLROAVKNESPVHIFLKPNSNEIG---KVESASPE-----DVR--- 119
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 69 APLKRPIDESAVRNFL---NEVAVFTGDLAATVSPQSGISOPGLNSGGCVPDARPPW 124
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 120 -YEFTILTKEVKGOTNKLASVTPVATINSLFNOI--KNQSCGISTASSPCITTRYPVDG 176
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 125 QYF-----YCCVPYASGRFCONEIKCKDNCGRNA-----D 156
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 177 CYAAHKMKROI-----LMNGGYDCEKO-----EYCGNLKAS--TGCCVAMS- 216
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 157 CYVANHOLNLCIKPGYIARARNGRCDDMKVQACSGDPHYTYTDLGRDYGTCTPPYFSQ 216
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 217 -----YHVA-----ILVSYKN-----ASGVTEKR 235
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 217 PCTTLPAPLWYVRAKNELPGKGYHISOVSEVEDLHNLTHVGRSKTALVNGV---O 273
 Qy 236 IIDPSLFS-----SGP-----VTDTANRACV-----NTSCG 262
 Db 274 VLTPEVFPKMTWTVAVRSGSTFTIENDOGVVFVFTTNSLCVQPDIPENFAGATTLCG 333
 Qy 263 SA-----SVSSAANTAGVY-----YRPSN-----SLYDNNLI----- 292
 Db 334 LAGNIDGKKLDDVYNNKNSGLAIKSSROPENNHNADEMKTEEDTWIDKFLILRPGOENCI 393
 Qy 293 -----NTNCVLTFRSLSGCS 308
 Db 394 NGOTLDNNTNCVSTISLSAQSCA 416

RESULT 2

AB3232
 conserved hypothetical protein PA3310 [Imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: AB3232
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lam,
 J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 gen
 A:Reference number: AB2950; MUID:20437337
 A:Accession: AB3232
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-551 <STO>
 A:Cross-References: GB:AE004753; GB:AE004091; NID:99949433; PIDN:AAG06698.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA3310
 C:Superfamily: Escherichia coli yidB protein

Query Match 6.0%; Score 101.5; DB 2; Length 551;
 Best Local Similarity 20.1%; Pred. No. 4.4;
 Matches 64; Conservative 40; Mismatches 95; Indels 119; Gaps 15;

Qy 58 GIKVSMFLAQFY-----IKPTKENEDYIGMLRAVKNESPV---HIFL 100
 Db 163 GCLLIYAVVSGSFYADIFREEDKLTHTINPT---NYIYAVGVYAAQRLVKEHLAV 218
 Qy 101 KPNSEIKGVESASPEDVRYFTILTEVKQTNKLASYIPVATLSL----- 149
 Db 219 EP-----IGDARQAPRALQROKSKLLVFVGEIAR-----ADHFSLNGYAREINPELSKO 269
 Qy 150 -----FNQIKNSCGSTA--SSPCITFRYVDCG---YARAHK-----MRQILMNN 191
 Db 270 DIVNFPQVR--SCGTSTAVSPCMFSQYPREDSDKAKATHEGLDILQAGVQVLMLEN 327
 Qy 192 GYDCEKQFYVGNLKAATGTC-----CVAMSYHAILVSYKNASG 230
 Db 328 NSDCK-----GTCLEVPNRDIPKTPSPFCDKNCLDESLVGL---QEIYDG 372
 Qy 231 VTEKRIT-----DPSLFSSGPVDTAMRNACVNTSCGSASVSYANTAGNVYRSPS 282
 Db 373 LQDDAIVLHSDGSHGPEYERYPRKMERFQVPCRTNQLGSCSKELVN----- 421
 Qy 283 NSLYDNNLINTNCVLTG 300
 Db 422 ---VYDNTILYTDHFLTK 436

RESULT 3
 S50820
 surface protein type 51B - Paramesium tetraurelia
 C:Species: Paramesium tetraurelia
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
 C:Accession: S50820

R:Scott, J.; Leeck, C.; Forney, J.
 Nucleic Acids Res. 22, 5079-5084, 1994
 A:Title: Analysis of the micronuclear B type surface protein gene in Paramesium tetra
 urelia
 A:Reference number: S50820; MUID:95098630
 A:Accession: S50820
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2395 <SCO>
 A:Cross-References: EMBL:U07603; NID:94672226; PIDN:AAA81947.1; PID:9467227
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
 C:Genetics:
 A:Genetic code: SCS5
 A:introns: 472/3; 1310/3; 1821/3
 C:Superfamily: G surface protein

Query Match 6.0%; Score 101; DB 1; Length 2395;
 Best Local Similarity 22.3%; Pred. No. 36;
 Matches 62; Conservative 22; Mismatches 94; Indels 100; Gaps 12;

Qy 122 KTLTEKVGQTNKILASVIPDVATLSLNFQIKNSCGSTASSPCITFRYPVD----- 175
 Db 464 KTCANAPSTNTNDLCVAFSLSCVSTVSTNAGCVDKTCENSLAQITC-----DKDLNKK 516
 Qy 176 -----GCTARAHKMQIIMNNGCYCEKQPYVGNLKAATGTCV-----AMSYHAIL 222
 Db 517 ACIMKGCYKRECVLSTTATHTADQTYDV--GCTLSNITGCVPLPKCEAITTEAACN 575
 Qy 223 VSYKNASGV-----TEKRIIDPSLFSSG----- 245
 Db 576 IRLQYTSVGVKSYPLCGWSSCIDKACSTAPKTTATTSQGTYSKCVANNPNVNSIQGC 635
 Qy 246 ---PYTDFA-----WR---NACVNTSCGSASVSYANTAGNVYRSP 281
 Db 636 QDLPTTCARKSTENCETIRTGPTCLMNSATSAQVCKSCSTASVTT---TGFELVFESN 692
 Qy 282 SNSLYDNNLINTNCVLTFRSLSGCSPAPDVSQC 319
 Db 693 TNCLAY---LSNSACTAN--NTADGCIPKP---SSCG 721

RESULT 4

Qy 124 GIBPT4
 gene 12 protein - phage T4
 N:Alternate names: tail fiber protein
 C:Species: phage T4
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 30-Jun-1993
 C:Accession: S01889
 R:Selivanov, N.A.; Prillipov, A.G.; Mesyanzhinov, V.V.
 Nucleic Acids Res. 16, 2354, 1988
 A:Title: Nucleotide and deduced amino acid sequence of bacteriophage T4 gene 12.
 A:Reference number: S01889; MUID:88189824
 A:Accession: S01889
 A:Molecule type: DNA
 A:Residues: 1-517 <SEL>
 A:Cross-References: EMBL:X06792
 A>Note: the authors translated the codon CAG for residue 279 as His
 C:Genetics:
 A:Gene: 12
 C:Superfamily: phage T4 gene 12 protein
 C:Keywords: tail fiber

Query Match 5.9%; Score 99.5; DB 1; Length 517;
 Best Local Similarity 20.8%; Pred. No. 5.8;
 Matches 77; Conservative 45; Mismatches 134; Indels 115; Gaps 15;

Qy 24 NGQENIGREKLSVNDKSLDKFGK-----TVP-----VGIDENGMIRKVSFML 66
 Db 79 NATEFYVGLTRSTNDEALAGVNNESSTPAKFYVALNNAFRTVSTESSNGYIKISLP 138
 Qy 67 TA-----QFEYKPTKENQYIGMLRQAVKNESPVHIFLKPMSNENKGVESASPEDVYR-- 119

Db 139 QALAGADDTAMTPLKTOOLAIKLAQIAPSETTA-----TESDOGVVQATVAQVROG 192
 Oy 120 -----YFKTILTKEVK-----QNKLASIYPVATLN-----SLFNQ 152
 Db 193 TLREGYAIPTFMNSSSTEYKGIKGTQSEVNSNNAVAVGATLNGSGSTTSMKV 252
 Oy 153 IKNOSCGTSTAASPCITFRYPVDCGYARAHKMQILMNNGYDCEKOFYVGNLKAS----- 207
 Db 253 VKL-----TTTAGS-----QSGGDASSALAMNADVIQORG-----QILYGLRIEDFTI 298
 Oy 208 -----TGTCVAMSYHVALVSKYKNASGYTEKRIIDPS-----LFSSGPTVTDAMR 253
 Db 299 ANGANGITG-----VRMTGGYIGGNRIVTQNEIDRTIPVGAIMMADSLPSDAMR 350
 Oy 254 NACVNTSCGSASVSANTAGNVYRSPNSGYLD-----NNLINTMCV 297
 Db 351 -FCHGGYASADCLYASRTIGRTGGNPSNPGLPDMGLFVGRSGRSHLTNPVNGNDQ 409
 Oy 298 LTKFSLSGCS 308
 Db 410 FGKPRLVGCT 420

RESULT 5

T39903
 serine-rich protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000

C:Accession: T39903

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, November 1998

A:Reference number: 221889

A:Accession: T39903

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-534 <L>N>

A:Cross-references: EMBL:AL033534; PIDN:CAA22127.1; GSPDB:GN00067; SPDB:SPBC215.13

A:Experimental source: strain 972h-; cosmid c215

C:Genetics:

A:Gene: SPDB:SPBC215.13

A:Map position: 2

C:Superfamily: pig submaxillary mucin

Query Match 5.9%; Score 99.5; DB 2; Length 534;

Best Local Similarity 22.7%; Pred. No. 6.1;

Matches 53; Conservative 39; Mismatches 123; Indels 25; Gaps 6;

Oy 85 MLROAVKNESPVHIFLKPNSEIGKVESASPEDVRYFKTILTKEVKQTNKLASVDPVA 144
 Db 24 LVHIOYGNSSFTKISLNLQEGDSQELQROEIRYGR--AAETGCTPTYYGVATPTSS 81
 Oy 145 TLNGLFNQIKNOSCGTSTAASPCITFRYPVDCGYARAHKMQILMNNGYDCEKOFYVGNL 204
 Db 82 SEPSLFSESAPPS--ETNYSPPSVSSYSDPATS-----QLPSTSFSEPTSESYPS 131
 Oy 205 KASGTCCVAMSYHVALVSKYKNASGYTEKRIIDPSLFSSGPTVTDAMRNACVNTSCSA 264
 Db 132 STESSLLDPPSSVSAIIPSS-----STVEVSTISSSLSDPLNSTSSLSSTSSSQP 187
 Oy 265 SVSS-YANTAGNVYRSPNSGYLDNNLI-----NTNCVLTFFSLSGCSPSPADPV 315
 Db 188 SVSSTSSSTFSSAAPTSTSSSVSSSVSSSSSSSSSTLTSSISTSSIPETSSSS 247
 Oy 316 SS 317
 Db 248 SS 249

RESULT 6

I64138

adhesin homolog HI1732 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Oct-1997
 C:Accession: I64138
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirsch, E.F.; Kerlavage
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Feldman
 , D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630
 A:Accession: I64138
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-298 <TIGR>
 A:Cross-references: GB:U32846; GB:U42023; NID:g1574588; PID:g1574589; TIGR:HI1732

Query Match 5.9%; Score 99; DB 2; Length 298;

Best Local Similarity 22.8%; Pred. No. 3;

Matches 66; Conservative 39; Mismatches 100; Indels 84; Gaps 15;

Oy 11 FVTY-LTFNSCADSNGNGEINGKELSYNDKIDFGKTVPGIDEENGMIKVSFMLTAQ 69
 Db 56 FVKVOSTEDIEDSAIKDDKNQALKAGDTLTLKAGNKLAKLDQ--GKSVTFAL-AK 112
 Oy 70 FYEIKPTKENBOY-IGMLROAVKNESPVHIFLKPNSEIGKVESASPEDVRYFKTILTK 128
 Db 113 DLDVKTAKVSPDLTIGNTTPAAGATP-----KVSITSTAD-----GKLAKG 155
 Oy 129 VKGQT---NKLASVDPVATLNSLFNQIKNOSCGTSTAASPCITFRYPVDCGYARAHK 184
 Db 156 TNGDTAVHLNGLASTLPDVT-----TNTGASTSVFSS-PSDIEKTRAATI 199
 Oy 185 RQILMNNGYDCEKOFYVGNLNASGTCCVAMSYHVALVSKYKNASGV-EKRIIDPSLFS 243
 Db 200 KDVL-NAGMNIK-----GAVVAGN-----TENDLVAGIDNEEFTIGDNTLDVVLTA 247
 Oy 244 --SGPVTDTAMRNACVNTSCGSASVSANTAGNVYRSPNSGYLDNN 290
 Db 248 KENGTTEVKF-----TPTSTVAKDNN 269

RESULT 7

JQ1792

Salp17R protein - vaccinia virus

N:Alternate names: Salp17R 64.7K protein

C:Species: vaccinia virus

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-Oct-2001

C:Accession: JQ1792; C38550

R:Smith, G.L.; Chan, Y.S.; Howard, S.T.

J. Gen. Virol. 72, 1349-1376, 1991

A:Title: Nucleotide sequence of 42kbp of vaccinia virus strain WR from near the right

A:Reference number: JQ1767; MUID:9125063

A:Accession: JQ1792

A:Molecule type: DNA

A:Residues: 1-564 <SMI>

A:Cross-references: DDBJ:D11079; NID:9222717; PIDN:BA01828.1; PID:9222743

A:Experimental source: strain WR

R:Howard, S.T.; Chan, Y.S.; Smith, G.L.

Virol. 180, 633-647, 1991

A:Title: Vaccinia virus homologues of the Shope fibroma virus inverted terminal repea

A:Reference number: A38550; MUID:91111982

A:Accession: C38550

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-564 <HOW>

A:Cross-references: GB:M58052

A:Experimental source: strain WR

C:Superfamily: A55R protein; POZ domain homology

F:7-110/Domain: POZ domain homology <POZ>

Query Match 5.9%; Score 99; DB 2; Length 564;

Best Local Similarity 22.1%; Pred. No. 7.2;

QY 277 YRSPNSSTLYDNLLNLTNCVLTKESSLG-CSPSPAP 313
 Db 734 YSRPNSNA-----SFLMSPHCTPSDISIVASSFASPTQP 767

RESULT 14

C42523
 A:55R protein - vaccinia virus (strain Copenhagen)
 C:Species: vaccinia virus
 A:Note: host Homo sapiens (man)
 C:Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 22-Oct-2001
 C:Accession: C42523
 R:Johnson, G.P.
 submitted to GenBank, June 1990
 A:Reference number: A31172
 A:Accession: C42523
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-564 <COH>
 C:Superfamily: A55R protein; PO2 domain homology
 F:7-110/Domain: PO2 domain homology <PO2>

Query Match 5.7%; Score 95; DB 2; Length 564;

Best Local Similarity 22.1%; Pred. No. 15; Matches 73; Conservative 49; Mismatches 110; Indels 98; Gaps 18;

QY 11 FVTVLTFNSCADSNGNOEIN--GKEKLSVNDSKLQFGKTPVPGIDEENGMIKVSFML-T 67
 Db 45 YSILFSNNFISN-EYEVNLSHLDQSYND--LIDYITGIPSLTNDN---VKIILST 97
 QY 68 AOFYEI-KPTKENEYIGMLRQAVKNESPVHIFLKPSNEIGKVESASPEDVRYEFTIL- 125
 Db 98 ADPLQIGSAITECENYI-LKNLCKNCIDFYIADKYNNK--KIESAS-----FNTILQ 148
 QY 126 --TKKVGQTNKLASYIPVATLNSLFNOIKN-----OSCGTS----- 161
 Db 149 NTLRLINDENFYLTESMIKITLSDMLINKEDFAPLILIKLESTOOSCTVELLCLR 208
 QY 162 -TASSP-----CINFRYPV--DCGVARAHKMR--QILMNGYDCERQ 198
 Db 209 ISLISPOVTKSLYSHRLVSYIECTIFLNNIAFLDESFPYHSIELISIGISNRDKISI 268
 QY 199 FVYGNLKASTGTCVAMSYVAIIVSYKNASGVTEKRIIDPSLFSSGPYTDFTAMNACVN 258
 Db 269 NCY-NHKNTWEMISRRYRCSPAVA-----VLDNIIYMMGGYDQSPYRSSKV- 315
 QY 259 TSCGSASVSSTANTAGNYRRSPNSLYLD 288
 Db 316 -----IAYNTCTNSMIVD 328

RESULT 15

A37023
 IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenza

N:Alternate names: Immunoglobulin A1 protease type 1

C:Species: Haemophilus influenzae

C:Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000

C:Accession: A37023

R:Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Killian, M.

Infect. Immun. 57, 3097-3105, 1989

A:Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemophilus influenzae

A:Reference number: A37023; MUID:89379374

A:Accession: A37023

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1541 <POU>

A:Cross-references: GB:X64357; NID:9433560; PIDN:CA45708.1; PID:g43561

A:Experimental source: serotype b

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloprotease

Query Match 5.7%; Score 95; DB 2; Length 1541;
 Best Local Similarity 23.5%; Pred. No. 59;
 Matches 65; Conservative 39; Mismatches 101; Indels 72; Gaps 17;

QY 17 FNSCADSNGNOELNGKEKLSVNDSKLQFGKTPVPGIDEENGMIKVSFMLTAOFYIKPT 76
 Db 37 FRDFAENKRFVSY-GATNVLVXDKNNKDLGTALP-----NGIPMIDFSVDVDKRI-AT 88
 QY 77 KENEYIGMLRQAVKNESPVHIFLKPSNEIGKVES--ASPEDVRYF----- 121
 Db 89 LNPQIVGVGKHYKNSGVSELHFGNLMGNMNN-GNAKAHADVSEENKRTFSVEKNEIPTKL 147
 QY 122 -KTLTFKEVKGQTNK-----LASVLPVATLNSLFNOIKNQSCGSTASPCITE 170
 Db 148 NKTVTED--QTKRRREDYWPRLDKFVTEVAPIEA-----STASSDAGTY 192
 QY 171 ----RTPVDGCTARAHKMQIILMNGYD-----CEKQFYGNLKAAGTCCVAMSYVA-- 220
 Db 193 NDQNKYP--AFVRLGSGSQFYTKKGDNYSLILNHEVGGNNLKLVD--AVYGIAGT 246
 QY 221 -ILVSYKN---ASGVTEKRIIDP-SLFSSGPYTDFA 251
 Db 247 PYKVHNNGLIGFGNSKEHSDPKGILSODPLTNYA 283

Search completed: June 27, 2002, 21:49:27
 Job time: 5215 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:59:19 ; Search time 39.56 Seconds
(without alignments)
313.202 Million cell updates/sec

Title: US-09-727-769A-8

Perfect score: 1679
Sequence: 1 MKNLFSLMMAFYTLTFNSC.....FSLISGCSPPADVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	5.9	527	1 VG12_BPT4	P10930 bacteriophage
2	99	5.9	564	1 VA55_VACCV	P24768 vaccinia vi
3	96.5	5.7	490	1 Y032_BORBU	O51063 borrelia bu
4	95.5	5.7	1679	1 YMF9_YEAST	O04958 saccharomyc
5	95	5.7	564	1 VA55_VACCV	P21073 vaccinia vi
6	95	5.7	1541	1 TGA1_HAETN	P42782 haemophilus
7	94.5	5.6	622	1 YF54_METUA	O38949 methanococc
8	94.5	5.6	850	1 D7-DICDI	O54654 dictyosteli
9	93	5.5	1183	1 CNA_STAAU	O54654 dictyosteli
10	92	5.5	344	1 CYS5_DICDI	P54640 dictyosteli
11	91.5	5.4	851	1 NID1_YEAST	P32336 saccharomyc
12	91.5	5.4	3130	1 DPOZ_HUMAN	O60673 homo sapien
13	91	5.4	761	1 NCAL_HUMAN	P13592 homo sapien
14	91	5.4	848	1 NCAL_HUMAN	P41951 homo sapien
15	89	5.3	1895	1 YLK3_CAEEL	O09550 caenorhabdi
16	88.5	5.3	1251	1 YOUS_CAEEL	P48849 hansenula w
17	88	5.2	386	1 RMAR_HANWI	P2128 drosophila
18	88	5.2	458	1 ME31_DROME	O92264 mus musculu
19	87.5	5.2	669	1 MTR1_MOUSE	O92264 mus musculu
20	87.5	5.2	928	1 PMP9_CHLPN	O92264 mus musculu
21	87	5.2	817	1 EMP1_CATRO	P35016 catharanthu
22	86.5	5.2	351	1 MREC_HAEIN	P44475 haemophilus
23	86.5	5.2	657	1 RES2_SCHPO	P41412 schistosach
24	86.5	5.2	821	1 VOIG_ECOLI	P76655 escherichia
25	86	5.1	383	1 FMA_FORGI	P13793 porphyromon
26	86	5.1	888	1 LAGC_DICDI	P42523 dictyosteli
27	85.5	5.1	638	1 GHR_RABIT	P19941 oryctolagus
28	85.5	5.1	1082	1 SP23_YEAST	P52210 saccharomyc
29	85.5	5.1	2715	1 G156_PARPR	P13837 paramycium
30	84.5	5.0	443	1 TOLB_RICPR	O92264 mus musculu
31	84.5	5.0	666	1 YEAT_YEAST	O92264 mus musculu
32	84.5	5.0	690	1 TRFE_ORYLA	P79819 oryzias lat
33	84.5	5.0	1401	1 LATA_LATMA	P23631 latrodectus

34	84.5	5.0	1620	1 DNM1_MOUSE	P13864 mus musculu
35	84.5	5.0	2704	1 G168_PARPR	P17053 paramycium
36	84	5.0	533	1 CAG2_MOUSE	O09200 mus musculu
37	84	5.0	725	1 NCAL_MOUSE	P13594 mus musculu
38	84	5.0	1115	1 NCAL_MOUSE	P13594 mus musculu
39	84	5.0	1694	1 TGA1_HAETN	P44969 haemophilus
40	83.5	5.0	405	1 ASSY_THBAC	O9hkfl thermoplasm
41	83.5	5.0	537	1 VGLF_PVM	P35949 pneumonia v
42	83.5	5.0	594	1 C1K1_YEAST	O01649 saccharomyc
43	83.5	5.0	775	1 UH87_HSV7J	P52363 human herpe
44	83	4.9	431	1 YG84_METTH	O27719 methanobact
45	83	4.9	435	1 FU27_USDMA	O99078 usiliago ma

ALIGNMENTS

```

RESULT 1
VG12_BPT4 STANDARD: PRT: 527 AA.
AC P10930:
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Short tail fiber protein (Protein Gp12) (p12).
GN 12.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D:
RX MEDLINE=86189824; PubMed=3357780;
RA Sellivanov N.A., Prilipov A.G., Mesyanzhinov V.V.;
RT "Nucleotide and deduced amino acid sequence of bacteriophage T4 gene
RL 12."
RN [2]
RP Nucleic Acids Res. 16:2334-2334(1988).
RA Kuter E., Aizaka F., Kunisawa T., Tsugita A., Morigi G.,
RA Mesyanzhinov V.V., Ruger W., Sliedham T., Thomas E.;
RT "Bacteriophage T4 genome analysis."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-57 FROM N.A.
RC STRAIN=D:
RX MEDLINE=89356257; PubMed=2548819;
RA Barrett B.K., Bergel P.B.;
RT "Using transposon Tn5 insertions to sequence bacteriophage T4 gene
RL 11."
RN [4]
RP DNA 8:287-295(1989).
RA SEQUENCE OF 1-4 FROM N.A.
RC STRAIN=D:
RX MEDLINE=89263746; PubMed=2726468;
RA Prilipov A.G., Sellivanov N.A., Efimov V.P., Marusch E.I.,
RA Mesyanzhinov V.V.;
RT "Nucleotide sequences of bacteriophage T4 genes 9, 10 and 11."
RL Nucleic Acids Res. 17:3303-3303(1989).
-1- FUNCTION: STRUCTURAL COMPONENT OF THE SHORT TAIL FIBER
(BASEPAIRS).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X06792; CAA29951.1; -
EMBL; AF158101; AAD42417.1; -

```

DR EMBL: M26253; AAA32495.1; -
 DR EMBL: X14192; CAA32398.1; -
 DR PIR: S01889; GIBPT4.
 DR PIR: C32479; C32479.
 KW Fiber protein.
 FT CONFLICT 33 33 H -> O (IN REF. 3).
 FT SEQUENCE 527 AA; 56214 MW; 808FF05DC3DAB7D CRC64;

Query Match 5.9%; Score 99.5; DB 1; Length 527;
 Best Local Similarity 20.8%; Pred. No. 1.6; Indels 115; Gaps 15;
 Matches 77; Conservative 45; Mismatches 134;

QY 24 NGNOEINGEKLSVNDKLFDEK-----TVP-----VGIDENGIKVSFML 66
 DB 89 NATEFYVGLTRYSTNDEALIGVNNESSITPAKFTVALINNAFETRVSTESNGVYKISLSP 148
 QY 67 TA-----QFEIKPTKENEDYIGMLQAVKNESPVHIFLKPSNEIGKVSASPEVYR-- 119
 DB 149 QALAGADDTTAMTPKTOALAIKLIQIAPSETTA-----TESDGVVOLATVAOVROG 202
 QY 120 -----YFKTILTKEVAG-----QTNKLASVDPVATLN-----SLFNQ 152
 DB 203 TLREGYAISPFTFMNSSSTEEYKGVKLGQSEVSNMNSVAATGATLNGRSTTSMRGV 262
 QY 153 IKNGSGCTSTASSPCITFRYPVDCYARAHKMKROIIMNNGYDCEKQFYVGNLKAS----- 207
 DB 263 VKL-----TTTACG-----OSGDSASSALAMNADVIOQRG-----QIYGTLRIDEDFTI 308
 QY 208 -----TGCCVAMSYHVAIIVSYKNASGVTEKRIIDS-----LSSSGVYTTAMR 253
 DB 309 ANGANITGT-----VRMTGCTYOGNRTVTONEDIRTPVGAIMMAADSLPSDAMR 360
 QY 254 MNCVNTSGSASVSYANTAGNYYRSPNSLYD-----NNLINTNCY 297
 DB 361 -FCHGTVASADCPVLASHTGTRYGNGPNPNGLPDMRGLEFVRSGSGSHLTNNVANGNDQ 419
 QY 298 LTRFSLSCS 308
 DB 420 FGRPRLGVGCT 430

RESULT 2
 ID VAS5_VACCV STANDARD; PRT; 564 AA.
 AC P24768;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein A55.
 GN A55R OR SALE17R OR SALEP21R.
 OS Vaccinia virus (strain WR).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OC NCBI_Taxid=10254;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91259063; PubMed-2045793;
 RA "Smith G.L., Chan Y.S., Howard S.T.;
 RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
 RT the right inverted terminal repeat."
 RL J. Gen. Virol. 72:1349-1376(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91111982; PubMed-1846491;
 RA "Howard S.T., Chan Y.S., Smith G.L.;
 RT "Vaccinia virus homologues of the Shope fibroma virus inverted
 RT terminal repeat proteins and a discontinuous ORF related to the tumor
 RT necrosis factor receptor family."
 RL Virology 180:633-647(1991).
 CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE POXVIRUSES KELCH FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: D11079; BAA01828.1; -
 DR EMBL: M58054; AAA48340.1; -
 DR PIR: C38550; C38550.
 DR PIR: JQ1792; JQ1792.
 DR InterPro: IPR000210; BTB, POZ.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF01344; BTB, 1.
 DR Pfam: PF01344; Kelch, 6.
 DR PRINTS: PR00501; KELCHREPEAT.
 DR SMART: SM00225; BTB; 1.
 DR PROSITE: PS50097; BTB; 1.
 KW Early protein; Repeat.
 FT DOMAIN 21 88
 FT REPEAT 252 297 BTB.
 FT REPEAT 298 346 KELCH 1.
 FT REPEAT 347 395 KELCH 2.
 FT REPEAT 397 441 KELCH 3.
 FT REPEAT 442 492 KELCH 4.
 FT REPEAT 494 539 KELCH 5.
 FT REPEAT 564 64690 KELCH 6.
 SQ SEQUENCE 564 AA; 64690 MW; FDAC2001FEEF7507 CRC64;

Query Match 5.9%; Score 99; DB 1; Length 564;
 Best Local Similarity 22.1%; Pred. No. 1.9; Indels 98; Gaps 18;
 Matches 73; Conservative 50; Mismatches 109;

QY 11 FYTVLTFPNSCADSNGNOEIN--GKEKLVNDKLFDEKFTVPYGVDEENGIKVSFML-T 67
 DB 45 YFSILFSPNNFIDSN-EYEVNLSHLDYOSVND-LIDYIGIPISLTNDN-----VKYILST 97
 QY 68 ADFEYI-KPTKENEYIGMLQAVKNESPVHIFLKPSNEIGKVSASPEVYRFTIL- 125
 DB 98 ADFLOIGSAITCECENYI-IKNLCSKNCIDFYADKYNNK-KIESA-----FWTILQ 148
 QY 126 -TKEVKQTNKLASVTPVATLNSLFNOIKN-----OSGGS----- 161
 DB 149 NTLRLINDENFYVLPESNMIKLSDMLNKNEDEFAPLLIKWLESTQSCVYELLRCR 208
 QY 162 -TASSP-----CITFRYPV--DGYARAHKMR-QILMNGYDCEKO 198
 DB 209 ISLSPQVTKSLYSHQVSSIVECTIFLNNIAFLDESFPYRHSIELISIGISNSHDKISI 268
 QY 199 FYVGNLKASTGTCVAMSYHVAIIVSYKNASGVTEKRIIDPSLFSGSPVTDAMRACVN 258
 DB 269 NCV-NHKRTWEMISRRRCFPAVA-----VLDNIITMGGYDQSPYNSKV- 315
 QY 259 TSCGSASVSYANTAGNYYRSPNSLYD 288
 DB 316 -----IAYNCTNSMIYD 328

RESULT 3
 ID Y032_BORBU STANDARD; PRT; 490 AA.
 AC O51063;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein BB0032.
 GN BB0032.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OC NCBI_Taxid=139;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gynn M.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kierlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Utterback T., Matthey L., McDonald L., Artach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
 burgdorferi.";
 RT Nature 390:580-586(1997).
 RL -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001117; AAC66427.1; -.
 DR TIGR: BB0032; -.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 27 47 POTENTIAL.
 SQ SEQUENCE 490 AA; 57590 MW; 0A4FD696A27E5548 CRC64;

Query Match 5.7%; Score 96.5; DB 1; Length 490;
 Best Local Similarity 26.0%; Pred. No. 2.5;
 Matches 50; Conservative 31; Mismatches 70; Indels 41; Gaps 10;

QY 5 FLSMAAFVTVITFNSCADSNGNOEINGKEKL-----SVNDSKLKDFGKTVPVGIDEE-NGM 59
 DB 309 FLIDBOVTVKXYEDFLKNEPKMALNNKELIKPOLVDENKFNQ---IGLNFAITGI 365
 QY 60 IKVSEFMLTAAQFEYK-PT-----KENQYIGMLRQAVKNSPVHIFLKPSNNEIGK 109
 DB 366 SYFAIEVANNYSKKLPFGFAKRLPISEWELV-----QKEPKNPLVI-----NEISK 414
 QY 110 -----VESASPEDVRYFTKTLITKEVKGQTNKLASVIPLEVATLNSLENOIKNOSCGTS- 161
 DB 415 KVGFWNLMOSSNFMEIAIFKN--EKNFYSNSNFYSILTEIRTSQGNNNLNLSTKASF 472
 QY 162 --TASSPCITFR 171
 DB 473 LKNMSSPNIGFR 484

RESULT 4
 YMF9_YEAST STANDARD; PRT; 1679 AA.
 ID YMF9_YEAST
 AC Q04958;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 187.1 kDa protein in OGCI-CNA2 intergenic region.
 GN YML059C OR YM9598.03C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Devlin K., Churcher C., Barrell B.G., Rajandream M.A.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SMS) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z46729; CAAB6716.1; -.
 DR SGD: S0004524; YML059C.
 DR InterPro: IPR001423; UPF0028.
 DR InterPro: IPR000595; cNMP-binding.
 DR Pfam: PF00027; cNMP-binding; 2.
 DR Pfam: PF01173; UPF0028; 1.
 DR SMART: SM01100; cNMP; 1.
 DR PROSITE: PS00042; cNMP_BINDING_3; 2.
 DR PROSITE: PS01237; UPF0028; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 104 124 POTENTIAL.
 SQ SEQUENCE 1679 AA; 187132 MW; 550FFCD4ACAF8E25 CRC64;

Query Match 5.7%; Score 95.5; DB 1; Length 1679;
 Best Local Similarity 22.8%; Pred. No. 14;
 Matches 77; Conservative 42; Mismatches 116; Indels 103; Gaps 16;

QY 17 FNSCADSNGNOEINGKEKLSVNDKLDKDFGKTVPVGIDEEGMKIVSFLTAAQFEIKPT 76
 DB 492 FKTYDKSSGSADLEKPKNSNNAASKLKPKPKAKP-----SGDID-----SLK-- 534
 QY 77 KENQYIGMLRQAVKNSPVHIFLKPSNNE-----IGKVESASPEDVRYFTKTLITKE 128
 DB 535 -----IANANANTSSNSLSLKPEFTHHPSRHVVLGSDQGNPD-----NOHK 613
 QY 129 VKGQTNKLASVIR-----DVATLNSLEFNOIKNOSCGTSTASSPCTTFPRYPDGCYARAHK 183
 DB 575 -----LLSNVPLSRTMDILSPNFIHNNRNKSGINTST5-----NOHK 613
 QY 184 -MQGLMNGYDCEKQF-----VYGNLKASTGTCV-AMSYHVALIVYKNASGVTEKR- 235
 DB 614 RSSRSSNNASVHSKFESSLPELRNAQLSTPSLDVTSVHDHHPVHLKGVRSRPR 673
 QY 236 -IDPSLFSSG--PYDTAMRNAC-----VNTSGSASVSSYANTA-----GNV 276
 DB 674 NLPTTSFAAQEETEDSALRMALVEAMLTGLGVKNSMVSSSSIAMSSLSNPOLNEM 733
 QY 277 YRSPNSLYLDNNLINTNCVILTKFSLSG--CSPEPAP 313
 DB 734 YSRPSNA-----SFLMSPHCPTSDISVASFSASPOTOP 767

RESULT 5
 VA55_VACC STANDARD; PRT; 564 AA.
 ID VA55_VACC
 AC P21073;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein A55.
 GN A55R.
 OS Vaccinia virus (strain Copenhagen).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10249;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
 RA Paoletti E.;
 RT "The complete DNA sequence of vaccinia virus.";
 RL Virology 179:247-266(1990).
 CC [2]
 CC COMPLETE GENOME.
 RP Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,

Db 193 NDONKP---AFVRLGSSQFTYKKGDVYSLILNHNHGVNKLKVG---AYTGIAGT 246
QY 221 -ILVSYKN---ASGVTEKRIIDP-SLFSSGPVDTA 251
Db 247 PYKNHNHNNGLIGFNGNSEHSDPKGILISDPLINYA 283

RESULT 7
YF54_METUA
ID YF54_METUA STANDARD; PRT; 622 AA.
AC Q58949;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M1554 precursor.
GN M1554.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=9637199; PubMed=6688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.S., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Colton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO A.FULGIDUS AF0817.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67596; AAB99574.1; -.
DR TIGR: M1554; -.
KW Hypothetical protein; signal; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 622 HYPOTHEICAL PROTEIN M1554.
SQ SEQUENCE 622 AA; 72282 MW; 86CA62F3854D73F CRC64;

Query Match 5.6%; Score 94.5; DB 1; Length 622;
Best Local Similarity 20.9%; Pred. No. 4.9;
Matches 73; Conservative 47; Mismatches 107; Indels 123; Gaps 20;

QY 197 KQVYGNLKAATGTCVAMSVHVALIVSYKNAAGTEKRIIDPSLFSSGPVDTAMRNAC 256
Db 206 -PIYWNANKICVD-----KYYIPEL-----PPIYSMD--FDYIIISR 240

QY 257 VNTSCGS-----ASVSYANTAGNVYRSPNSLYIDNNLINTNCVLAKF 301
Db 241 INIKSGKVENSIATVGNKKT---LYMSKNILYFAVNLKINEKKLMNF 286

RESULT 8
D7_DICTD
ID D7_DICTD STANDARD; PRT; 850 AA.
AC P54682;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CAMP-inducible prepsore protein D7 precursor.
GN D7.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX3;
RX MEDLINE=95080502; PubMed=7988791;
RA Agarwal A., Sloger M.S., Oyama M., Blumberg D.D.;
RT "Analysis of a novel cyclic Amp inducible prepsore gene in
RT Dictyostelium discoideum: evidence for different patterns of camp
RT regulation.";
RL Differentiation 57:151-162(1994).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PREPSORE CELLS.
CC -1- INDUCTION: BY CAMP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U25143; AAA73514.1; -.
DR Dictydb; DD02038; -.
KW Sporulation; signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 850 CAMP-INDUCIBLE PREPSORE PROTEIN D7.
FT DOMAIN 470 475 POLY-GLN.
FT DOMAIN 555 568 POLY-ASN.
FT DOMAIN 728 738 POLY-GLN.
SQ SEQUENCE 850 AA; 95343 MW; 13BA634CCE7AA502 CRC64;

Query Match 5.6%; Score 94.5; DB 1; Length 850;
Best Local Similarity 23.1%; Pred. No. 7.3;
Matches 74; Conservative 42; Mismatches 111; Indels 93; Gaps 15;

QY 10 AFVYVLTN---SCADSNQNOEINK---EKLVSVDKLRPGKVPVPGIDEENGIYVS 63
Db 8 SFVYVLLISCTISTSTQNOISEDGKAJNKLNEIKEVED---VYVOFENAVNV- 60

QY 64 FMLTQAEYETKPTKEVYIG---MLRQAVKNE---SPVHFLKPNSEIKGVESA 113
Db 61 -----IEELAVENORQELLEGVKHDLSRTNTYINGLDLKIQSYLPKDNKKSKVVEA 114

QY 114 SPEDVRYFKTILTEKVKQJTNKLASVDPV--ATLNSLFNOIK-----NQSCGST- 162
Db 115 FS-----SQONNIGSSIGDSTGASTSPQFOSINGLSGASGSGSTGTG 160

QY 163 -----ASSPCTIFRPVPGCCARAHKMQIILMNGCYDDEKQFYGNLKA--TGICCVAMS 216
Db 161 DSDSKTNEALITFSKSVSTTDROESITIGOVAIT-----AKDSIGATTGTGLGVSST 211

```

QY 217 YVAVILVSKNASCVTETKRIIDPSLFSSGPTVDTAMRNACVNTSCGSASY----- 266
DB 212 ARVGGQITNGRAGQ-----QVITGDNITGVGRGAVTTASAVANTVGEFLGSSRTG 262
QY 267 -SSYANTAGNVYRSPSNXY 285
DB 263 GSSSACTVGNVT-----SDSY 278

RESULT 9
CNA_STAAU STANDARD: PRT: 1183 AA.
AC 053654:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FDA 574;
RX MEDLINE=92163839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin.",
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RL Lindberg M., Hoecek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN-FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoecek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus.",
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoecek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin.",
RL Nat. Struct. Biol. 4:833-838(1997).
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, M81736; AAA20874.1; .
CC PDB, 1AMX; 24-JUN-98.
CC InterPro: IPR001899; Gram_pos_anchor.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
CC Signal: Repeat; Transmembrane; Cell wall; 3D-structure.
CC SIGNAL 1 29 POTENTIAL.

```

```

FT CHAIN 30 1183 COLLAGEN ADHESIN.
FT DOMAIN 30 1157 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1158 1177 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1178 1183 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 151 318 COLLAGEN-BINDING.
FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 1151 1156 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT REPEAT 533 719 B1.
FT REPEAT 720 906 B2.
FT REPEAT 907 1093 B3.
SQ SEQUENCE 1183 AA; 133066 MW; 6A1CC072E575D76 CRC64;

Query Match 5.5%; Score 93; DB 1; Length 1183;
Best Local Similarity 21.4%; Pred. No. 15;
Matches 73; Conservative 39; Mismatches 127; Indels 102; Gaps 16;

QY 2 KNLFSMAFVTVLT-----FNSCADSNGNOETNGKE-----KLSVNDSKLDEGK-T 48
DB 3 KNV-LKFWFIMLNTITPLFNK-----NEAFARDISTVNTDLTVSPSKIEDGKTT 55
QY 49 VYVGDDEENG-----MIVSEMLTQFTEIKPTKENEQYIGLRQAVNESPVHIFLRP 102
DB 56 VKMTFDDKNGKIQNGDMIVAMPETSG-----TVKIEGSKTVPLTVRGEQVGAVIHP 108
QY 103 NSNEI---GKVSASPEDRYRFPKTLITKEVKG---QTKKLASVDPVATLNSLFN-QIK 154
DB 109 DQATITFNDKVEKLS--DVSGFAEF---EVQGNLTQTITSDKAVATTISGKNSITVYH 163
QY 155 NOSCGTSTASSPCITFRFYPDGCYARAHKMQIILNNGYDCEKQFYVYGLKASTGCVA 214
DB 164 KSEAGTSSV-----FYV-----KTGDMLEP 183
QY 215 NSYHAILVSYKNASCVTETKRIIDPSLFSSGPTVDTAMRNACY-----NTSCGSASVSY 269
DB 184 DTHVRFMLINNNKSYSKDITIKDQIOGGQQLDLSLTIINVTGTHSNVYSGQSAITDF 243
QY 270 A-----NTAGNVYRSPSNXYLYDNMLINTNCVLT 299
DB 244 EKAFRSKTIVNTKNTITVITPQIGSYNSFSINKTKIT 284

RESULT 10
CYS5_DICDI STANDARD: PRT: 344 AA.
ID CYS5_DICDI
AC P54640:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cysteine proteinase 5 precursor (EC 3.4.22.-).
GN CPRE OR CP5.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4;
RX MEDLINE=96081966; PubMed=7499424;
RA Soua G.M., Hirai J., Menta D.P., Freeze H.H.;
RT "Identification of two novel Dictyostelium discoideum cysteine
RT proteinases that carry N-acetylglycosamine-1-P-modification.",
RL J. Biol. Chem. 270:28938-28945(1995).
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN THE VEGETATIVE PHASE AND DECREASES
CC WITH THE START DEVELOPMENT. REAPPEARS IN LOW LEVELS WHEN THE
CC FRUITING BODY IS FORMED.
CC -1- PTM: PHOSPHOGLYCOSYLATED, CONTAINS GLCNAc-ALPHA-1-P-SER RESIDUES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC PAPAIN FAMILY OF THIOL PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L36205; AAA92018.1; -
 DR HSSP: P07711; 1CJL.
 DR MEROPS: C01.081; -
 DR DictyDD: DD01061; cPRE.
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; Thiolprot_act_site.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: PR00705; PAPA1N.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; 1.
 DR PROSITE: PS00640; THIOI_PROTEASE_ASN; 1.
 KW Hydrolyase; Thiol protease; Lysosome; Zymogen; Glycoprotein;
 KW Phosphorylation; Signal.
 FT SIGNAL 1 17
 FT PROPE 18 111 POTENTIAL.
 FT CHAIN 112 344 ACTIVATION PEPTIDE (POTENTIAL).
 FT DOMAIN 196 340 CYSTEINE PROTEINASE 5.
 FT ACT_SITE 136 136 SER-RICH.
 FT ACT_SITE 272 272 BY SIMILARITY.
 FT ACT_SITE 311 311 BY SIMILARITY.
 FT ACT_SITE 311 311 BY SIMILARITY.
 FT DISULFID 133 174 BY SIMILARITY.
 FT DISULFID 167 207 BY SIMILARITY.
 FT DISULFID 265 333 BY SIMILARITY.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 344 AA; 37212 MW; 82F3F58BC147BA8 CRC64;

Query Match 5.5%; Score 92; DB 1; Length 344;
 Best Local Similarity 22.7%; Pred. No. 3.7;
 Matches 76; Conservative 36; Mismatches 117; Indels 106; Gaps 16;

QY 55 EENGIKIKSEMLTAOFYIKPTKEKNEYIGMLROA-VKNESPVHFLKPNSEIKGVSA 113
 Db 44 EEFGARVNIETANDMYVOOMNSKSEYVLGNFADITNEEYRNLYCT-----KPDAS 97
 QY 114 SPEVDARYKTLTKEKVGOTKLA-----SVIPDVATLNSLFNOIKNO-SCG---- 159
 Db 98 S-----LGTQDEKVTHTSSAASKMRSEGAVT-----VKMGCGCGGWS 138
 QY 160 -TSTAAS-----PCITFRYPVDG--CYARAHKMRQILMNGYDC 195
 Db 139 FSTGTGTEGAHFQSKGELVSLSEONLIDCSTENSCDGLMTYA---FEYIINNNGIDT 194
 QY 196 EKQFYGLKAKSTGTCCVAMSYHVALIVSYKNAAGVTETKRI-----IDPS--- 240
 Db 195 ESSYPT--KANGKCEKXSENSGATLSSYKTYTAGSESSLESAAVNPVSAVAIDASHQS 251
 QY 241 --LSSGVTDTAMRNAC-----VNTSCGSASVSYANTAGNYVYRSPSY 285
 Db 252 FQLTSG-----IYEPRESSENLDHGVLAAGVGGSGSSGSSQSSGGLN-SSASSNEY 306
 QY 286 LYDNNLINTNCVLTKFSLLSCSPSPADVSSCGF 320
 Db 307 WIKNSMGTSMGIEGYILMSRNRDNNCGIASSASF 341

RESULT 11
 NDDI_YEAST
 ID NDDI_YEAST STANDARD; PRT; 851 AA.
 AC P32336; Q08895;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NDDI protein.
 GN NDDI OR YOR373W.

OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CX NCBI_TaxID=4932;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Dulic V., Zanolari B., Riezman H.;
 RL Submitted (SEP-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Delius H., Hebling U., Hofmann B.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -i- FUNCTION: REQUIRED FOR NUCLEAR DIVISION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X62147; CA444073.1; -
 DR EMBL: Z75281; CA99704.1; -
 DR PIR: S19056; S19056.
 DR SGD: S0005900; NDDI.
 DR InterPro: IPR001611; LRR.
 DR Pfam: PF00560; LRR; 5.
 DR PRINTS: PR00019; LEURICHRPT.
 KW Cell cycle.
 FT DOMAIN 223 239 POLY-ASN.
 FT DOMAIN 266 274 POLY-SER.
 FT CONFLICT 5 6 TO -> SE (TN REF. 2).
 SO SEQUENCE 851 AA; 94103 MW; CB9F040863C1315 CRC64;

Query Match 5.4%; Score 91.5; DB 1; Length 851;
 Best Local Similarity 21.2%; Pred. No. 13;
 Matches 34; Conservative 26; Mismatches 57; Indels 43; Gaps 6;

QY 24 NNGOENKEREKLVNDSKLDGKTPVGVIDEENGMIVSEMLTAQFYIKPTKEKNEYI 83
 Db 314 NQGLDVSSSHSLDNTSSNQSFATWVPTGDNHTNG--KAPSLDKKAVELSTPGD--V 369
 QY 84 GMLQAVKNESPVHFLKPNSEIKGVSAPEVDARYKTLTK-----EVKGGQNKLAS 138
 Db 370 GYRKQKIQDEENL-----ANSDD-----TPLDPKPNDELFTYKNGTRAKVGGQRTSRS 417
 QY 139 V-----IPDVATLNSLFNOIKNO 156
 Db 418 ISNSNLEAHKKLKTFFPAERYEDITISSEVNTSPNETEKQ 457

RESULT 12
 DPOZ_HUMAN
 ID DPOZ_HUMAN STANDARD; PRT; 3130 AA.
 AC O6673; O43214;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE DNA polymerase zeta catalytic subunit (EC 2.7.7.7) (hREV3).
 GN REV3L OR POLZ OR REV3.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RP [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Fetal brain;
 RC MEDLINE=98284025; PubMed=9618506;
 RA Gibbs P.E.M., McGregor W.G., Maher V.M., Nisson P., Lawrence C.W.;
 RT "A human homolog of the Saccharomyces cerevisiae REV3 gene, which
 RT encodes the catalytic subunit of DNA polymerase zeta.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:6876-6880(1998).
 RM [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Bone marrow, and leukocyte;
 RX MEDLINE-99202263; PubMed-10102035;
 RA Lin W., Wu X., Wang Z.,
 RT "A full-length cDNA of hREV3 is predicted to encode DNA polymerase
 RL zeta for damage-induced mutagenesis in humans.";
 RM Mutat. Res. 433:89-98(1999).
 RP [3]
 RP SEQUENCE FROM N.A.
 RA Murakumo Y., Rasio D., Roth T., Negriini M., Croce C.M., Fishel R.;
 RT Cloning and characterization of hREV3, the human homolog of S.
 RL cerevisiae REV3.";
 RM Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RP [4]
 RP SEQUENCE OF 79-3130 FROM N.A.
 RA Roth T., Rasio D., Murakumo Y., Negriini M., Croce C.M., Fishel R.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 RM [5]
 RP SEQUENCE OF 79-3130 FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE-99126315; PubMed-9925914;
 RA Morelli C., Mungall A.J., Negriini M., Barbanti-Brodano G., Croce C.M.;
 RT "Alternative splicing, genomic structure, and fine chromosome
 RL localization of REV3L.";
 CC Cytogenet. Cell Genet. 83:18-20(1998).
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + (DNA)(N).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.
 CC -1- DOMAIN: ITS C-TERMINAL PART COULD SERVE AS THE CATALYTIC DOMAIN
 CC DURING NUCLEOTIDE POLYMERIZATION, WHILE ITS N-TERMINAL PART COULD
 CC PROVIDE SITES FOR PROTEIN-PROTEIN INTERACTIONS WITH OTHER FACTORS
 CC DURING TRANSLATION DNA SYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: AF058701; AAC24357.1; -
 DR EMBL: AF071798; AAC24009.1; -
 DR EMBL: AF157476; AAP40184.1; -
 DR EMBL: AF179428; AAG09402.1; -
 DR EMBL: AF179429; AAG09403.1; -
 DR EMBL: AF035537; AAB88486.1; -
 DR EMBL: AF078695; AAC28460.1; -
 DR MIM: 602776; -
 DR InterPro: IPR02064; DNA_POL_B.
 DR Pfam: PF00136; DNA_POL_B_1.
 DR Pfam: PF03104; DNA_POL_B_exo; 2.
 DR PRINTS: PR00106; DNAPOLB.
 DR SMART: SM00486; POLBc; 1.
 DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
 KM Transferrase; DNA-directed DNA polymerase; DNA replication;
 KM DNA-binding; DNA repair; Nuclear protein; Zinc-finger; Polymorphism.
 FT ZN-FING 3042 3057
 FT C4-TYPE (POTENTIAL).
 FT ZN-FING 3086 3104
 FT C4-TYPE (POTENTIAL).
 FT Q -> H.
 FT /FTID=VAR_008516.
 FT VARIANT 389 389
 FT S -> T.
 FT /FTID=VAR_008517.
 FT VARIANT 1540 1540
 FT K -> E.
 FT /FTID=VAR_008518.
 FT VARIANT 2607 2607
 FT S -> T.
 FT /FTID=VAR_008519.
 FT CONFLICT 237 237
 FT E -> Q (IN REF. 4 AND 5).
 FT CONFLICT 1156 1156
 FT Y -> C (IN REF. 4 AND 5).

SQ SEQUENCE 3130 AA; 352782 MW; 1C0700900F10BB14 CRC64;
 Query Match
 Best Local Similarity 19.0%; Score 91.5; DB 1; Length 3130;
 Best Local Similarity 19.0%; Pred. No. 67;
 Matches 60; Conservative 51; Mismatches 112; Indels 93; Gaps 11;
 2 KNLF5MAFVYVLP-----NSCADSNGNGQENKGRKLSV--DSKLEDFGKTPV 50
 DB 608 KGIDNSVTFTNESTYSMKYPGSLSTVSHNSKNSKKEILLPVSCSSIFDYEDIP 667
 QY 51 VGIDENGMKIVSEMLTAQFYELIKPTKENEDYIGLRQAQVNSPYHFLKPNSEIGV 110
 DB 668 -----SVTRQVPSSKRYTN-----IRKIEKSPFIHMRHPNENTIGK- 704
 QY 111 ESASPDVRYFKTILRKE--VKGQTFKLASVIP----- 141
 DB 705 NSPFSDDLHNSKKNVSEENKGNSTALSLPSSSTENCCELLSCGENTVHVSINSTA 764
 QY 142 DVATLNSL-----FMQIKNSGCTSPASSPCTTFRRPV--DCGYARAHKROIIMN---- 190
 DB 765 DESGLNKLTIRYEEFQEHKTEKPSLSQAQAHMFFPSVLSNCLTRPKLSPTTYTLQPG 824
 QY 191 ---NGYDCEKQFYVGNLKAAGTCGVAWSYHVAIIIVSYNAGVTEKRIIDPSLFSSGTV 247
 DB 825 NKPSRLKLNKRKLAGEHQTSTKS-----SETGSKDNFIQNNPCNSNDE 868
 QY 248 TDTAMRNACVNTSCGS 263
 DB 869 KDNALASDLTKTTRGA 884
 RESULT 13
 NC2_HUMAN
 ID NC2_HUMAN STANDARD: PRT: 761 AA.
 AC P13592; P13593;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neutral cell adhesion molecule, phosphatidylinositol-linked isoform
 DE Precursor (N-CAM 120) (NCAM-120) (CD56 antigen).
 GN NCAM1 OR NCAM.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (N-CAM 120).
 RC TISSUE-Skeletal muscle;
 RX MEDLINE-89305258; PubMed-3253057;
 RA Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,
 RA Elsom V., Moore S.E., Goridis C., Walsh F.S.;
 RT "Complete sequence and in vitro expression of a tissue-specific
 RT phosphatidylinositol-linked N-CAM isoform from skeletal muscle.";
 RL Development 104:165-173(1988).
 RN [2]
 RP SEQUENCE OF 491-761 FROM N.A. (N-CAM 120).
 RC TISSUE-Skeletal muscle;
 RX MEDLINE-87301755; PubMed-2887295;
 RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
 RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
 RT "Human muscle neural cell adhesion molecule (N-CAM): identification
 RT of a muscle-specific sequence in the extracellular domain.";
 RL Cell 50:1119-1130(1987).
 RN [3]
 RP SEQUENCE OF 491-655 FROM N.A. (SECRETED ISOFORM).
 RX MEDLINE-8907552; PubMed-3203385;
 RA Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
 RA Dickson G., Walsh F.S.;
 RT "Alternative splicing generates a secreted form of N-CAM in muscle
 RT and brain.";
 RL Cell 55:955-964(1988).
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN

```

CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD56 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd56.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16841; CAA34739.1; -.
DR EMBL: M17409; AAA59912.1; -.
DR EMBL: M22094; AAA59910.1; -.
DR EMBL: M22092; AAA59911.1; -.
DR EMBL: M22091; AAA59911.1; JOINED.
DR PIR: S07784; IJHUNG.
DR PIR: A31635; A31635.
DR MIM: I16930; -.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003598; IG_C2.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGC2; 5.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW GPI-anchor; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 761
FT FT
FT DOMAIN 34 103
FT FT
FT DOMAIN 132 196
FT FT
FT DOMAIN 228 294
FT FT
FT DOMAIN 322 392
FT FT
FT DOMAIN 419 486
FT FT
FT DOMAIN 518 595
FT FT
FT DOMAIN 660 727
FT FT
FT DISULFID 41 96
FT FT
FT DISULFID 139 189
FT FT
FT DISULFID 235 287
FT FT
FT DISULFID 329 385
FT FT
FT DISULFID 426 479
FT FT
FT CARBOHYD 222 222
FT FT
FT CARBOHYD 315 315
FT FT
FT CARBOHYD 347 347
FT FT
FT CARBOHYD 423 423
FT FT
FT CARBOHYD 449 449
FT FT
FT CARBOHYD 478 478
FT FT
FT VASPLIC 635 655
FT FT
FT SEQUENCE 761 AA: 83770 MW: FQCAD3292D7AB67E CRC64;
SO

```

```

Oy 156 OSCGTASSPCITPR-----YPVDG-CYARAH-----KMRQILMNNGDCE 196
Db 328 -TCEASGPIPSITWRTSTRNISSEKTLIDGMVYRSHARVSSLTKSIQTIDAGEICT 386
Oy 197 KQFYGNLKASTGTCVAMSVHVALVSKNAGVTEKRIIDPSLFSGPYTDAMRNAC 256
Db 387 ASNTIGQDSQS-----MYLEVQYA-----PKL--GGPVAVYTWENQ 421
Oy 257 VNTSC-----GSASVSY-----ANTAGVYRSPNSLT-----YDNNLINTNCL 298
Db 422 VNTCEVAPYSATISWRDGLPSSNYSNIKINYTPSASYLEVTPSEDFGNCTA 481
Oy 299 T-----KPSLISGCSPPAPDV 315
Db 482 VNRIGQESLEFLVQADTPS-SPSI 505

RESULT 14
NCAL_HUMAN
ID NCAL_HUMAN STANDARD: PRT: 848 AA.
AC P13591; Q16180; Q15829;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140) (CD56 antigen).
GN NCAM1 OR NCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94356433; PubMed=8075973;
RA Saito S., Tanih Y., Tachibana I., Hayashi S., Kishimoto T., Kawase I.;
RT "Complementary DNA sequence encoding the major neural cell adhesion
RT molecule isoform in a human small cell lung cancer cell line.";
RL Lung Cancer 10:307-318(1994).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=91250739; PubMed=1710251;
RA Lanier L.L., Chang C., Azuma M., Ruitenberg J.J., Hemperly J.J.,
RA Phillips J.H.;
RT "Molecular and functional analysis of human natural killer cell-
RT associated neural cell adhesion molecule (N-CAM/CD56).";
RL J. Immunol. 146:4421-4426(1991).
[3]
RN SEQUENCE OF 491-848 FROM N.A.
RX MEDLINE=87301755; PubMed=2887295;
RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA Moore S.E., Cox R.D., Quinn C., Pult W., Walsh F.S.;
RT "Human muscle neural cell adhesion molecule (N-CAM): identification
RT of a muscle-specific sequence in the extracellular domain.";
RL Cell 50:1119-1130(1987).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD56 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd56.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```


Qy 308 SPSPAPDVSSC 318
| : |
Db 1289 STNOVCISNQC 1299

Search completed: June 27, 2002, 21:59:22
Job time: 665 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2002, 21:58:26 ; Search time 102.74 Seconds
(without alignments)
538.820 Million cell updates/sec

Title: US-09-727-769a-8
Perfect score: 1679
Sequence: 1 MKNFLSMMAFVTLTFNSC.....FSLSGCSPSPAPDVSSCGF 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp._archaea:*
2: sp._bacteria:*
3: sp._fungi:*
4: sp._human:*
5: sp._invertebrate:*
6: sp._mammal:*
7: sp._mhc:*
8: sp._organelle:*
9: sp._phage:*
10: sp._plant:*
11: sp._protist:*
12: sp._virus:*
13: sp._vertebrate:*
14: sp._unclassified:*
15: sp._virus:*
16: sp._dactariap:*
17: sp._archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1679	100.0	320	2 Q9A008	Q9ag8 chryseobact
2	108.5	6.5	342	13 P70004	P70004 xenopus lae
3	108	6.4	927	5 002364	002364 caenorhabdi
4	102	6.1	2673	4 096SC3	096sc3 homo sapien
5	102	6.1	5636	4 096RW7	096rw7 homo sapien
6	101.5	6.0	551	16 09HYT4	09hyt4 pseudomonas
7	101	6.0	2395	5 027167	027167 paramecium
8	100	6.0	848	15 040366	040366 human immun
9	99.5	5.9	534	3 094317	094317 schizosacch
10	99	5.9	594	10 09S746	09s746 bacillidopsi
11	98.5	5.9	419	16 09KB36	09kb36 bacillus ha
12	97	5.8	534	17 026459	026459 methanother
13	96.5	5.7	564	12 072736	072736 cowpox viru
14	96.5	5.7	260	16 099R23	099r23 staphylococ
15	96.5	5.7	5458	5 090459	090459 plasmodium
16	96	5.7	423	2 09F961	09f961 anaplasma m

17	95.5	5.7	673	16 097TK7	097tk7 clostridium
18	95	5.7	423	2 09Z620	09z620 anaplasma m
19	93.5	5.6	1894	11 P70206	P70206 mus musculu
20	93	5.5	347	2 093R80	093r80 porphyromon
21	93	5.5	594	10 09SX23	09sx23 arabidopsis
22	92	5.5	1154	10 09LP06	09lp06 arabidopsis
23	91.5	5.4	427	17 0971P3	0971p3 sulfobus
24	91.5	5.4	2013	16 092EK2	092ek2 listeria in
25	91.5	5.4	2909	4 09NU25	09nu25 homo sapien
26	91	5.4	392	10 09FZJ9	09fzj9 arabidopsis
27	91	5.4	396	10 09ZW53	09zw53 arabidopsis
28	91	5.4	876	5 09XZM5	09xzm5 mya arenari
29	90.5	5.4	843	5 002264	002264 caenorhabdi
30	90.5	5.4	1166	12 082666	082666 avian infec
31	90	5.4	243	5 018804	018804 caenorhabdi
32	89.5	5.3	1124	10 09SF08	09sf08 arabidopsi
33	89.5	5.3	1836	16 092E25	092e25 listeria in
34	89.5	5.3	1836	10 09LXK4	09lxk4 arabidopsi
35	89.5	5.3	2500	5 096223	096223 plasmodium
36	89.5	5.3	2510	5 094658	094658 plasmodium
37	89	5.3	595	9 09AZR8	09azr8 bacterioph
38	89	5.3	595	16 09CPS9	09cps9 laccococcus
39	89	5.3	672	5 095QP7	095qp7 caenorhabdi
40	89	5.3	1073	16 09A4M6	09a4m6 caulobacter
41	88.5	5.3	283	17 026933	026933 methanother
42	88.5	5.3	301	6 095WM0	095wm0 lepus capen
43	88.5	5.3	450	2 P74849	P74849 salmonella
44	88.5	5.3	466	16 0971L19	0971l19 clostridium
45	88.5	5.3	491	5 09XUP6	09xup6 caenorhabdi

ALIGNMENTS

RESULT	ID	Q9A008	PRELIMINARY:	PRT:	320 AA.
Q9A008	Q9A008				
AC	Q9A008				
DT	01-JUN-2001	(TRENBLREL. 17, Created)			
DT	01-JUN-2001	(TRENBLREL. 17, Last sequence update)			
DT	01-DEC-2001	(TRENBLREL. 19, Last annotation update)			
DE	PROTEIN-GLUTAMINASE PRECURSOR.				
GN	PROA.				
OS	Chryseobacterium proteolyticum.				
OC	Bacteria; CFB group; Flavobacteriia; Flavobacteriaceae;				
OX	Chryseobacterium.				
OC	NCBI_TaxID=118127;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-9670:				
RA	MEDLINE-21153247; PubMed-11231294;				
RA	Yamaguchi S., Jeenes D.J., Archer D.B.;				
RT	"Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme				
RT	that deamidates glutamyl residues in proteins: purification,				
RT	characterization and gene cloning.";				
RL	Eur. J. Biochem. 268:1410-1421(2001).				
DR	EMBL; AB046594; BAB21508.1; -.				
KW	Signal.				
FT	SIGNAL	1	21	POTENTIAL.	
FT	CHAIN	136	320		
SO	SEQUENCE	320 AA;	35044 MW;	C67823D2BC131410 CRC64;	
Query Match					
Best Local Similarity 100.0%; Score 1679; DB 2; Length 320;					
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MKNFLSMMAFVTLTFNSCADSNGNOENKELSVNDSKLPKGVPGIDEENMI	60		
Db	1	MKNFLSMMAFVTLTFNSCADSNGNOENKELSVNDSKLPKGVPGIDEENMI	60		
QY	61	KVSFMTAOFYFIKPTKENEOYIGMLRQAVKNESPVHFLFPLNSNEIGKVESAPEDVRY	120		

```

Db 61 KSFMLTAQFYETKPKNEQYIGMLRQAVKNESPVHILKPNSEIGKVESASPEDVRY 120
OY 121 FKTILTRKVGQTNKLASYIPDVATLNSLFNQIKNOSCGTSTASSPCITFRYPVDCYAR 180
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 FKTILTRKVGQTNKLASYIPDVATLNSLFNQIKNOSCGTSTASSPCITFRYPVDCYAR 180
OY 181 AKRMROIIMNNGYDCEKQVYGNLKAISTGTCVANSYHVALVSKYKNASGVTEKRIIDS 240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 AKRMROIIMNNGYDCEKQVYGNLKAISTGTCVANSYHVALVSKYKNASGVTEKRIIDS 240
OY 241 LSSSGPVDTARNACVNTSCGSASVSANTAGNYYSPNSLYDNKLINTNCVLTK 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 LSSSGPVDTARNACVNTSCGSASVSANTAGNYYSPNSLYDNKLINTNCVLTK 300
OY 301 FSLSGCSPSPAPDVSSCGF 320
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 FSLSGCSPSPAPDVSSCGF 320

RESULT 2
ID P70004 PRELIMINARY; PRT; 342 AA.
AC P70004;
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-FEB-1997 (TREMBLrel. 19, Last annotation update)
DE ALPBI-MICROGLOBULIN/BIKUNIN (AMB) PRECURSOR.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_Taxid=8353;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Kawahara A., Hikosaka A., Sasado T., Hirota K.;
RT "Thyroid hormone-dependent repression of alpha1-microglobulin/bikunin
  (AMB) gene expression during amphibian metamorphosis.";
RL Dev. Genes Evol. 206:355-362(1997).
DR EMBL; D87752; BAA13453.1; -.
DR HSSP; P02760; BIK.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR005566; Lipocalin_CytfABP.
DR Pfam; PF00014; Kunitz_BPTI; 2.
DR Pfam; PF00061; Lipocalin; 1.
DR PRINTS; PR00759; BASICPTASE.
DR PRINTS; PR00179; LIPOCALIN.
DR SMART; SM00131; KU; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00213; LIPOCALIN; 1.
KW Serine protease inhibitor; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 342 POTENTIAL.
SQ SEQUENCE 342 AA; 38558 MW; 3050508BFFC5C47B CRC64;

Query Match 6.5%; Score 108.5; DB 13; Length 342;
Best Local Similarity 23.1%; Pred. No. 1;
Matches 49; Conservative 30; Mismatches 58; Indels 75; Gaps 12;
OY 67 TAOFY-----EIKPTKENOYIGMLRQAVKNESPVHILKPNSEIGKVESASPEDV---- 118
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 TVKLYGRSPDLRPTLVDERQFALAOGIPEDS---IVMLPNNGE-----CSPGTEIYVR 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 119 -HYFTILTRKVGQTNKLASYIPDVATLNSLFNQIKNOSCGTSTASSPCIT----- 168
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 196 RTQRAVRVLPREEEGSGME-----NSPFSKNGKSGSCLAPASGCLGNHNRYFYN 244
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 169 -----FRYPVDCYARAHKRMROIIMNNGYDCEKQFYG-----NLKASTGTCCYA 214
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 SSTMACETFRQY--GGCLGN-----NNNFSEKECHLDCTEAACXLPITTPGCKTA 293

```

```

OY 215 ---WSYHVA-----ILVSYKNASG-----VTER 234
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 294 KTHWAFDAAGKCVTFESYGGCGGNGNQFYTER 325

RESULT 3
ID 002364 PRELIMINARY; PRT; 927 AA.
AC 002364; P90955; O62351;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 102.3 KDA PROTEIN T01D3.6 IN CHROMOSOME V.
GN T01D3.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A
RC STRAIN=BRISTOL N2;
RA Steward C.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM T01D3.6A (SHOWN HERE) AND
    -!- ISOFORM T01D3.6B; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; 281527; CAB04280.1; JOINED.
DR EMBL; 281110; CAB03262.1; -.
DR EMBL; 281527; CAB03262.1; JOINED.
DR EMBL; 281527; CAB04279.1; JOINED.
DR EMBL; 281110; CAB04279.1; JOINED.
DR EMBL; 281110; CAB03263.1; -.
DR EMBL; 281527; CAB03263.1; JOINED.
DR HSSP; P02671; IF2D.
DR WormRep; T01D3.6A; CE12964.
DR WormRep; T01D3.6B; CE18164.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR002919; TIL.
DR InterPro; IPR001846; Vwd.
DR Pfam; PF00147; Fibrinogen_C; 1.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00186; FBG; 1.
DR SMART; SM00216; VWD; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 1.
KW Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;
KW Hydroxylation; Hypothetical protein; Repeat.
FT VARSPLIC 405 416 MISSING (IN ISOFORM T01D3.6B).
SQ SEQUENCE 927 AA; 103413 MW; 71AEF0A61FC2B266 CRC64;

Query Match 6.4%; Score 108; DB 5; Length 927;
Best Local Similarity 19.2%; Pred. No. 3.8;
Matches 85; Conservative 47; Mismatches 101; Indels 210; Gaps 24;
OY 35 LSVNSKTLDFGKTVV-----GIDENGMIKVSFMLTAQFYETKPKNEQYI 83
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 15 LAAEDDLKCYGADCEVPYIMDTDECGSGLECES---KETLLAMISHYO---KEKHDEYL 68
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 84 GMLRQAVKNESPVHILKPNSEIG---KVESASPE-----DVR--- 119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 69 APLTKRIPRESAVRNRL-----NEVAVTGDLATFVSPQCSICSPGLCNSGGCVPARFWM 124
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 120 -YFKTILTRKVGQTNKLASYIPDVATLNSLFNQI--KNOSCGTSTASSPCITFRYPVDC 176

```

```

Db 125 QYF-----YCVCPYASGRFCONEIKCKDNKSGKNA-----D 156
QY 177 CYARAHKMRQI-----LMNNGYDCEKO-----FVGNLTKAS-TGTCCVAMS- 216
Db 157 CVYANHOLNICKPKGYTARRNGRDCDMKVQACMSGDHPYVYTGDLRPDYOGTCPPYFSQ 216
QY 217 -----YHVA-----ILVSYKN-----ASGVTEKR 235
Db 217 PCTLPAPYLMTSVRAKNELPBGKGYHISQVSEVEDLNLITHVDGRSKTALVNGV---Q 273
QY 236 IIDSLES-----SGP-----VTDTAMRNACV-----NTSCG 262
Db 274 VLTWPYFPNKNMTVTRVAFSGSTFTIENDOGVVVTFITYNSLCVQVDPDIPFENGATLTCG 333
QY 263 SA-----SVSSYANTAGNMY-----YRSPSN-----STLYDNMLI----- 292
Db 334 LAGNIDGKKLDVYVKNKSGSYLAIKSSRQPPENNHNADFKETEDTWTIDKFLILRPQGENCI 393
QY 293 -----NTNCLVLTFRFSLSGCS 308
Db 394 NGOTLDNNTNCVSTISILAQSCA 416

```

```

RESULT 4
Q96SC3 PRELIMINARY; PRT; 2673 AA.
AC 096SC3:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FIBULIN-6 (FRAGMENT).
GN FIBL-6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Kostka G., Jimpl R.;
RT "Partial sequence of fibulin-6 with a C-terminal region related to
RT domain II and III of the fibulin family."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ306906; CAC37630.1; .
FT NON_TER
SQ SEQUENCE 2673 AA; 291017 MW; BEAEC30B8340E272 CRC64;

```

```

Query Match 6.1%; Score 102; DB 4; Length 2673;
Best Local Similarity 19.8%; Pred. No. 43;
Matches 66; Conservative 58; Mismatches 125; Indels 84; Gaps 16;

```

```

QY 27 OEINGEKLSVNDKSLKDFGK-----TVPVGIDENGMIKY-----SFM 66
Db 614 QTLGGGEVLRISTAGVEDTGRYTCLASSPADDDKEYLVHRVHPNPIAGTDEPRDITVLR 673
QY 67 TAQF-YEIKPTKENQYIGMLRQAVKNESPVHI-----FLKNSNEIGKVESASPEDV 118
Db 674 NRQVTLCKSDAVPPVPTVTLRNGERLOATPRVRILSGRYLQIINNADLG-----DTA 726
QY 119 RYFETILTKVEKQGTNK-----LASVIPDV-----ATLNSLFNQINKQSGTSTASPCITE 170
Db 727 NY--TCVASNIAAGKTRREFILITVNPPIKGGPQSLVLLKSKYVLECIAGVPTPRITW 784
QY 171 RYPVDCYARAHKMRQIIMNNGYDCEKOFYVGNLKAISTGT-CVYAMSYHVALIVSYKNAS 229
Db 785 R-KDGAVALAGNHAARYLSLENGF-----LHIOAHVTDGRYLCA-----TMAA 827
QY 230 GYTEKRITD-----PSLFSGPVTDFTAMRNACVNTSCGSASVSYANTAGNVYRSPNS 284
Db 828 G-TDRRRIDLOVHVPPSIAPGPPTNMVTVLVNQTLLACEATGIP-----KPSIN 874

```

```

QY 285 YLYDNNLINTNCVLTKFSLSGCS---PSPAPD 314
Db 875 WRKNGHLNVQNONSYRLSSGSLVISPVD 907

```

```

RESULT 5
Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC 096RW7:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEMICENTIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RT "Human hemicentin gene."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF156100; AAK68690.1; .
SQ SEQUENCE 5636 AA; 613660 MW; F000B319CED7B52C CRC64;

```

```

Query Match 6.1%; Score 102; DB 4; Length 5636;
Best Local Similarity 19.8%; Pred. No. 11e+02;
Matches 66; Conservative 58; Mismatches 125; Indels 84; Gaps 16;

```

```

QY 27 OEINGEKLSVNDKSLKDFGK-----TVPVGIDENGMIKY-----SFM 66
Db 3577 QTLGGGEVLRISTAGVEDTGRYTCLASSPADDDKEYLVHRVHPNPIAGTDEPRDITVLR 3636
QY 67 TAQF-YEIKPTKENQYIGMLRQAVKNESPVHI-----FLKNSNEIGKVESASPEDV 118
Db 3637 NRQVTLCKSDAVPPVPTVTLRNGERLOATPRVRILSGRYLQIINNADLG-----DTA 3689
QY 119 RYFETILTKVEKQGTNK-----LASVIPDV-----ATLNSLFNQINKQSGTSTASPCITE 170
Db 3690 NY--TCVASNIAAGKTRREFILITVNPPIKGGPQSLVLLKSKYVLECIAGVPTPRITW 3747
QY 171 RYPVDCYARAHKMRQIIMNNGYDCEKOFYVGNLKAISTGT-CVYAMSYHVALIVSYKNAS 229
Db 3748 R-KDGAVALAGNHAARYLSLENGF-----LHIOAHVTDGRYLCA-----TMAA 3790
QY 230 GYTEKRITD-----PSLFSGPVTDFTAMRNACVNTSCGSASVSYANTAGNVYRSPNS 284
Db 3791 G-TDRRRIDLOVHVPPSIAPGPPTNMVTVLVNQTLLACEATGIP-----KPSIN 3837
QY 285 YLYDNNLINTNCVLTKFSLSGCS---PSPAPD 314
Db 3838 WRKNGHLNVQNONSYRLSSGSLVISPVD 3870

```

```

RESULT 6
Q9HYT4 PRELIMINARY; PRT; 551 AA.
AC 09HYT4:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE HYPOTHEICAL PROTEIN PA3310.
GN PA3310.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Slover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Watrener P.,

```

RA Hickey R.L., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Collier L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiser J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an
 opportunistic pathogen."
 RT Opportunistic pathogen.
 RL Nature 406:959-964(2000).
 DR EMBL: AE004753; AAG06698.1; .
 DR InterPro: IPR003371; DUF146.
 DR Pfam: PF02418; DUF146; 1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 551 AA; 61421 MW; 3E3EAF5C308BE815 CRC64;

Query Match 6.0%; Score 101.5; DB 16; Length 551;
 Best Local Similarity 20.1%; Pred. No. 6.5;
 Matches 64; Conservative 40; Mismatches 95; Indels 119; Gaps 15;

QY 58 GKIKSFMLTAQFE-----IKPTKENEDYIGMLQAVKNESFV--HIFL 100
 DB 163 GCLLYAVVWVGSEFYAFIFREEDKLTHFINPT---NYIYAVGKAKQRLVIKEHLVY 218
 QY 101 KNSNIEIGVESASPEDVAFYFTILKEVGOTNKLASYIPDVATINSL----- 149
 DB 219 EP-----IGDAAQAPALOROKKSLVYVGETAR-----ADHFSUNGARETNPELSKQ 269
 QY 150 ----FNQIKNQGCGTSTA--SSPCITFRYPVDCG---YARAHK-----MRQITLNN 191
 DB 270 DIVNFTQVH--SCGSTAVSPCMFSQYPREYSDKAKATHEGLDILORAGQVLMLEN 327
 QY 192 GTDCEKQFYGNLKAATGTC-----CYAMSYHVALIYKRNASG 230
 DB 328 NSDCK-----GTCLRVNPNRDIPTQSPSPCDKKNCLDESLVLGL---QEIYDG 372
 QY 231 VTEKRIT-----DPSLFSSGSPVDTFAMRNACVNTSGSASVSAYTAGNVYRSPS 282
 DB 373 LDDDAITIVHSGSHQPEYERKPKEMERFQYPCRTNQLGSCSKELVN----- 421
 QY 283 NSLYDNMLINTNCVLTFR 300
 DB 422 ---VYDNTILYTDHFLTK 436

RESULT 7
 Q27167 PRELIMINARY; PRT; 2395 AA.
 AC Q27167;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE 51B TYPE SURFACE PROTEIN.
 OS Parametium tetraurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Parametium.
 OX NCBI_TaxID=5888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-STOCK 51;
 RX MEDLINE=95098630; PubMed=7800503;
 RA Scott J., Leeck C., Forney J.;
 RT "Analysis of the micronuclear B type surface protein gene in
 Parametium tetraurelia."
 RT Nucleic Acids Res. 22:5079-5084(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-STOCK 51;
 RA Scott J.M., Leeck C.L., Forney J.D.;
 RL Genetics 133:189-198(1993).
 DR EMBL: U07603; AAA81947.1; .
 DR EMBL: L04795; AAA16710.1; .
 DR InterPro: IPR002895; Parametium_SA.
 DR InterPro: IPR003659; PSI.

DR InterPro: IPR001680; WD40.
 DR Pfam: PF01508; Parametium_SA; 28.
 DR SMART: SM00423; PSI; 3.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 SO SEQUENCE 2395 AA; 246430 MW; 8632E022649B2D61 CRC64;

Query Match 6.0%; Score 101; DB 5; Length 2395;
 Best Local Similarity 22.3%; Pred. No. 45;
 Matches 62; Conservative 22; Mismatches 94; Indels 100; Gaps 12;

QY 122 KTLTREVGGQTNKLASVLPDVATINSLNFQIKNQGCGTSTASPCITRYPVD----- 175
 DB 464 KTCANAPSTNTNDLCVAELSSCTVNSTAGCDVCKTCENSLSAQTIC-----DKDLNKK 516
 QY 176 -----GCYARAHKROILMNNNGYDCEKQFYVGNLASTGTCCV-----AMSYHVALI 222
 DB 517 ACIMKCKCYKRECVLASSTTATHADQYTDV--GCTLSNLTGCVPLPLCEAITTEAACN 575
 QY 223 VSYKNSGV-----TEKRIDPSLFSSG----- 245
 DB 576 IRLOQTVGKVSYPICGNGSSCIDKACSTAPKTTATTSQGYKSGCVANNPNVNGSIQCG 635
 QY 246 ---PVYDFA-----NR---MACVNTSGSASVSANTAGNVYRBP 281
 DB 636 QDLPTTCARAKSTENCEITRTGFPPTCLMNSATSCVCKSCSTASVYT---TTGFLTVFSN 692
 QY 282 SNSLYDNMLINTNCVLTFRSLSGSPSPADVSSCG 319
 DB 693 TNCILAV---LNSNACIAN--NRADGCIKPR-----SSCG 721

RESULT 8
 ID 040366 PRELIMINARY; PRT; 848 AA.
 AC 040366;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
 DE ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FRMR;
 RX MEDLINE=98216834; PubMed=9557756;
 RA Blanchard A., Ferris S., Chameret S., Guetard D., Montagnier L.;
 RT "Molecular evidence for nosocomial transmission of human
 immunodeficiency virus from a surgeon to one of his patients."
 RL J. Virol. 72:4537-4540(1998).
 DR EMBL: U85919; AAC59357.1; .
 DR InterPro: IPR000328; Env-GP41.
 DR InterPro: IPR000777; GP120.
 DR Pfam: PF00516; GP120; 1.
 DR Pfam: PF00517; GP41; 1.
 KW AIDS; Coat protein; Glycoprotein; Polypeptide; Transmembrane.
 FT NON_TER
 SO SEQUENCE 848 AA; 96100 MW; 4C224022C395F34B CRC64;

Query Match 6.0%; Score 100; DB 15; Length 848;
 Best Local Similarity 21.2%; Pred. No. 15;
 Matches 60; Conservative 39; Mismatches 94; Indels 90; Gaps 13;

QY 101 KNSNIEIGVESASPEDVAFYFT---ILTKYKGTNKLASYI--PDVATL----- 146
 DB 133 KNSN---TTESPPIEDPREMTNCSFNMTTLRKTKKVSLFYKLDVQDEENSNS 189
 QY 147 NSLFDIKNQGCGTSTASPC---ITFR--YPVDCYARAHKM----- 184
 DB 190 NGNYSYRLVNCNTSTTQACPKVSEFPIPIHYCAPAGYAILKCKDKNKGPCSNVST 249

```

QY 185 -----ROLNNGYDCEKOFYVGNLKASTGTCCVAMSYYHVALIVSYKNASGVT 232
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 250 VOCHTGKIPVYSTOLLNGLSLAEKIVIRSKNKSDSTSIIVQLVH-PVKINCFPNMNT 308
QY 233 EKR-----IIDPSLFSSGPGVTDPAAMNACVN-----TSC 261
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 309 KRSVRIKPGAFYTRDIIIDIRQAQSCYVNTKTEWRNTLLNVSEOLKEYPFNKKATFENRSSG 368
QY 262 GSASVSSYA-NTAGNVYRSPS---NSLYDNMNIINTNCVLT 299
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 369 GDEITTHSFNCRGEFFCYNTSGLFNNTLLHDTN--NTNGMT 409

RESULT 9
ID 094317 PRELIMINARY: PRT: 534 AA.
AC 094317;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHEICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
DE PRECURSOR.
GN SPKC215.13.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX Schizosaccharomyces.
NX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajadream M.A., Barrell B.G., Rieger M.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- SIMILARITY: TO YEAST STAL.
DR EMBL; AL013534; CAA22127.1; -.
KW Hypothetical protein; Glycoprotein; Signal.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 1 14 HYPOTHEICAL SERINE-RICH PROTEIN C215.13.
FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 426 426 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 534 AA; 53361 MW; D62DC15C80A698B4 CRC64;

Query Match 5.9%; Score 99.5; DB 3; Length 534;
Best Local Similarity 22.7%; Pred. No. 9;
Matches 55; Conservative 39; Mismatches 123; Indels 25; Gaps 6;
QY 85 MLROAVKNESPVIHFLKPNSEIGKVESASPEDVRYEFKTLITKEVKGQTNKLASYIPDVA 144
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 24 LVHQIIGSSFTKISLNLNLEGDSOELOKROEITRYCR--AAETGGIPPTYGYVTPPSS 81
QY 145 TLNLSLFNQIKNQSGCTSTASSPCITFRYPVDCYARAHKMQOILMNNGYDCEKOFVGNL 204
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 82 SEPSIFSESAPPS-ETNYSVSPVSYSIDPATNS-----QLPSTSEFFSPSTSEYTPS 131
QY 205 KASGTCTCVASYYHVALIVSYKNASGVTEKRLIDPSLFSSGPGVTDPAAMNACVNTSCSA 264
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 132 STESSSLDPSSVSAILPS---STVEVSISSSLSSDPLTSSPTSSSLSSSTSSSQP 187
QY 265 SVSS-YANTAGNVYRSPSNTLYDNMLI-----NTN-VLTFESLSCGSPSPADVD 315
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 188 SVSSSTSSSTFSSAAPTSTSSSYLSSSVSSSSSSPSSSSSTLSSSLSTSSIPSTSSS 247
QY 316 SS 317
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 248 SS 249

RESULT 10
ID 095746 PRELIMINARY: PRT: 594 AA.

```

```

AC 095746;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ACE (AT1G72970/F3N23_17).
GN ACE OR F3N23.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Araki T., Nakatani-Goto M.;
RT "ADHESION OF CALYX EDGES (ACE), cDNA.";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Federapfel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Alatafi H., Araujo R., Huizar L., Rowley D., Chen S., Hartman P.,
RA Hicks R., Huerta M., Mason S., Siepel J., Zimmerman M., Buehler E.,
RA Dunn P., Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J.,
RA Liu S., Lueros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carininci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamuya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinohara K., Davis R.W., Theologis A., Ecker J.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB027458; BAA77837.1; -.
DR EMBL; AC008017; AAD55644.1; -.
DR EMBL; AY054193; AAL06854.1; -.
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR001072; GMC_oxred.
DR Pfam; PF00732; GMC_oxred; 1.
DR PROSITE; PS00339; AA_trna_ligase_II_2; UNKNOWN_1.
DR PROSITE; PS00624; GMC_OXRED_2; UNKNOWN_1.
SQ SEQUENCE 594 AA; 65344 MW; D6A8682FBBF73881 CRC64;

Query Match 5.9%; Score 99; DB 10; Length 594;
Best Local Similarity 21.0%; Pred. No. 11;
Matches 49; Conservative 28; Mismatches 74; Indels 82; Gaps 11;
QY 81 QYIGMLROAVK-----NESP---VHIFLKPNSEIGKVESASPEDVRYEFKTLITKEV 129
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 353 QTVGITKMGVVEASTGCGSPESIHTHYGLMSKNEL-----FSTIPAKOR 399
QY 130 KGQTNKLASVLPDVALNSLEN-----QIKNQSGCTSTASSPCITFRY 172
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 400 RPEATQ-AVITRNKYLQLEAFNGSFLEKLAYPEISRGHLSLVN--TNVDNPSVTPNY 454
QY 173 ---PVD--GCVARAHKMQOILMNNGY-----DCEKOFVGNLKAST----- 208
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 455 FKHPVDLORCYEATRLVSKVYTSNRFLLWYTCQDKONVAKMISLAKANINLRPKOLNTK 514
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
QY 209 -----GTCCVAMSYYHVALIVS-----YKNASGVTEKRLIDPSLFSSGPGVTD 249
      ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|  ||:|
Db 515 SMAQFCRQDTVYTIMHYHGGCLVGVVSPNRYKLVGDRLRVIDGTFDSSPGFN 567

RESULT 11
ID 09KB36 PRELIMINARY: PRT: 419 AA.

```

AC 09K836:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE BH2099 PROTEIN.
 GN BH2099.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125/JCM 9153;
 RA MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL, AP001514; BAB05818.1;
 KW Complete proteome.
 SO SEQUENCE 419 AA; 46289 MW; 50B84D6C23EAE2C4 CRC64;

Query Match 5.9%; Score 98.5; DB 16; Length 419;
 Best Local Similarity 23.0%; Pred. No. 8;
 Matches 65; Conservative 38; Mismatches 99; Indels 81; Gaps 12;

OY 12 VVLPFNSGADNSNGNEIKGKESLVNDSKLDKDFCTVPVGIDEEGMKVSFMLTAQRY 71
 DB 30 VVLPFNSGADNSNGNEIKGKESLVNDSKLDKDFCTVPVGIDEEGMKVSFMLTAQRY 71
 OY 72 E-IKPTKENEQYIGMLRQAVKNESPVHIFLKPNSEIKGVESASPEDVRYFKTI----- 124
 DB 70 EKLIRKASFKETEVARLVQVGDIPILFLF-----EYLSESDIQDMITVEELNE 119
 OY 125 ----LTKREVAGQTNKLASVTPDVATLNSLFPNQIKNSCGTSTSPCITTRYPYDGCYARA 181
 DB 120 ABAVAAVKEGADGIL-VIPDGFVNSLYAL----VGEPTASS----LIYKEMETSNS 170
 OY 182 HKMKROL-----MNGGYCEKQFYVGNLKAISG-----TCCVANS 216
 DB 171 ATLQYIIDFLDMNNHFLQDMGEVQENAPEGGLEKTVGEGAFILTYITISMSALFA 230
 OY 217 YHVAILVSYKNASGVYTER--RII-----DPSLFSSGPTVTAW 252
 DB 231 LFLAATVATKGTGEIRQKVFHRIILLDRKPIILFVGKIVSTAF 273

RESULT 12
 O26459 PRELIMINARY; PRT; 534 AA.
 AC 026459:
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE CONSERVED PROTEIN.
 GN MTH359.
 OS Methanothermobacter thermautotrophicus.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacter.
 OX NCBI_TaxID=145262;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RA MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Harridge T., Bashirzaden R., Blakey D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Viscare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jivani N., Caruso A., Bush D., Saefer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Nollig J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 DR EMBL, AE000821; AAB84865.1;
 DR InterPro: IPR002931; Transglut_core.
 DR Pfam: PF01841; Transglut_core; 1.
 DR SMART: SM00460; TGC; 1.
 KW Complete proteome.
 SO SEQUENCE 534 AA; 57628 MW; 588412BD8D342E26 CRC64;

Query Match 5.8%; Score 97; DB 17; Length 534;
 Best Local Similarity 20.9%; Pred. No. 14;
 Matches 57; Conservative 34; Mismatches 108; Indels 74; Gaps 9;

OY 44 DFGKITPVGI-----DEENGMIVSEMLTQ-----FYEIKPTKENEQYIGMLQ 88
 DB 277 NCGKNTVPAPVATAGEAANDAGRGTLTRLELYLKVASGVQGEKKNRAPN--YAGSSKG 334
 OY 89 AVKNESPVHIFLKPNSEIKGVESASPEDVRYFKTITKEVGQGTNKLASVTPDVATLNS 148
 DB 335 RISTSELYV-----SISRLSFSSMKR-----LPNTVTVTR 366
 OY 149 LFNQIKNS-----CGTSTASSPCITFRYPVD-----GCYARAK----- 183
 DB 367 LASSLKNRPENDPYRNGEJARYLASASCVPDIRSLASEITRGLTSTFSAEAVFGM 426
 OY 184 MQQIIMNNGYDCEKQFVYVNCNLSKASTCCVANSYHAILVSYKNASGVYTERKILIDPSLFS 243
 DB 427 VHDNIIVSYFYNTKIGAVTTLKRNIGCV----DTHHLVALARAAGIPARYVHGTCNPT 482
 OY 244 SGPTVTAMRNACVNTSCGSASVSYANTAGNV 276
 DB 483 SGVYGHVWAQQLVGDPTWYAADATSSRNSLGVY 515

RESULT 13
 O72736 PRELIMINARY; PRT; 564 AA.
 AC 072736:
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE A54R PROTEIN.
 GN A54R.
 OS Cowpox virus (CPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10243;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GRI-90;
 RA MEDLINE=98229462; PubMed=9568042;
 RA Shechelkinov S.N., Saifonov P.F., Tolentin A.V., Petrov N.A.,
 RA Ryazankina O.I., Gutorov V.V., Kotwal G.J.;
 RT "Species-specific differences in genome organization of cowpox,
 RT smallpox, and vaccinia viruses.";
 RL Virology 243:432-460(1998).
 DR EMBL: Y15035; CAJ75274.1;
 DR InterPro: IPR000210; BTB_POZ.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF00651; BTB; 1.
 DR Pfam: PF01344; Kelch; 6.
 DR PRINTS: PR00501; KELCHREPEAT.
 DR SMART: SM00225; BTB; 1.
 DR PROSITE: PS00997; BTB; 1.
 SO SEQUENCE 564 AA; 64825 MW; EF268497F8DC6AFA CRC64;

Query Match 5.8%; Score 97; DB 12; Length 564;
 Best Local Similarity 21.8%; Pred. No. 15;
 Matches 72; Conservative 52; Mismatches 108; Indels 98; Gaps 18;


```

QY 11 FVTLTSCADSGNGOEIN--GREKLSVNDSKLKDFGKTVPGIDEENGMIKVSFML-T 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 45 YFSLFNSNFDN-EYEVNLSHLDYGVND--LIDYIYGIPLSLTNDN----VKYIIST 97
QY 68 AOFTEI-KPTENECYIGMLQAVKNESPVHIFLKPNSENGKVESASPEVRYFKTIL- 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 98 ADFQIGSAITECEKYI-LKMLCSNCDIFYIADKYNNK--KIESAS-----FNTILR 148
QY 126 --TKEVGQTKKLASVIDVATLNSLFNQIKN-----QSCGTS----- 161
DB 149 NIKRIINDENKRYLTEESMIKILSDMLINKNEDEPAPILIKWLESTQOCTVELLCRL 208
QY 162 -TASSP-----CITFRYPV---DGCYARAHKMR--QILMNGYDCEKQ 198
DB 209 ISLSLPQYKLSYHQVLSYIECTEFLNNIAPLDESFPFRHSTIELISIGISNSHDKISI 268
QY 199 FVYGNLKASTGCCVAMSYHVAIIVSYKNASGVTEKRIIDPSLSSGCVPTDPMRACVN 258
DB 269 NCY-NHKNTWEMISSRRYRCSFAVA-----VLNHIYWMGGYDQSPYKRSKY- 315
QY 259 TSCGSASVSYANTAGNYRSPNSLYD 288
DB 316 -----IAYNTCTNSWIYD 328

RESULT 14
099RZ3
ID 099RZ3 PRELIMINARY: PRT: 260 AA.
AC 099RZ3
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MODA PROTEIN.
GN MODA OR SA2074.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21311952; PubMed-11418146;
RA Kuroda M., Ohta T., Uchiiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Itoh T., Kanamori M.,
RA Matsunaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani U.Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hatakeyama H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003136; BAB43371.1;
DR HSSP: P37329; IMOD.
KW Complete proteome.
SQ SEQUENCE 260 AA; 29117 MW; 7A5D4A01A4482C4D CRC64;

Query Match 5.7%; Score 96.5; DB 16; Length 260;
Best Local Similarity 22.5%; Pred. No. 6.3;
Matches 42; Conservative 39; Mismatches 59; Indels 47; Gaps 9;
QY 1 MKNLFLSMAFVTVLTFNSCADSGNGOEINGEKLSVNDSKLKDF--GKTVPGIDEENG 58
DB 3 MKRFIAIYMALELYLA--GCSNSNDN-----NESKKDDADNGKQEIQVAAAS 49
QY 59 MIKVSFMLTAQF---YEIKPTKENQYIGMLQAVKNESPVHIFLKPNSENG--KVES 113
DB 50 LTDVTKKLASEPKKHKADIKFNFGSGALRKQIESCAPVDVFNVSAMTKVDALDKNK 109
QY 114 SPEVRYFKTIL-----TKEVGQTKKLASVTPDVATLNSLFNQIKN-----NSL 149
DB 110 AHDYTKYAKNSLVLLIGDKDSNYSYVDLK--DNDKLA--LGEVKTVPAGKYAKOYLDNNNL 166

```

```

QY 150 FNOIKNQ 156
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 167 FKEVESK 173

RESULT 15
090459
ID 090459 PRELIMINARY: PRT: 5458 AA.
AC 090459: 090PG4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ERYTHROCYTE MEMBRANE-ASSOCIATED GIANT PROTEIN ANTIGEN 332.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
RT "Molecular cloning and structure analysis of the Plasmodium falciparum
RT erythrocyte membrane-associated giant protein Ag332 (Pf332) gene."
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF202180; AAF15293.3;
DR InterPro: IPR001313; PUM.
SQ SEQUENCE 5458 AA; 615269 MW; 4F6A37D92BD80172 CRC64;

Query Match 5.7%; Score 96.5; DB 5; Length 5458;
Best Local Similarity 22.6%; Pred. No. 2.9e+02;
Matches 36; Conservative 32; Mismatches 64; Indels 27; Gaps 5;
QY 21 ADSNGOINCKEKLVSNDSKLKDFGKTVPGIDEE--NGMIKVSFMLTAQFYEI--KPT 76
DB 4654 AEVSYTEEIPGVETTNESHGKGVNVEYVVDNSYNDIEQFDDSSIEITYVDSKV 4713
QY 77 KENQYIGMLQAVKNESPVHIFLKPNSENGKVES-----ASPEDVR----- 119
DB 4714 FHKEENYDSFREFVRSDDNIHSRKNNEFVKIKIENKESGVHPTETYSKNIAEVPS 4773
QY 120 --YFKTITL---KEVGQTKKLASVTPDVATLNSLFNQ 152
DB 4774 HIMFKNYTEVEVPREVKYEENTVEVEFEVTSKENIIEE 4812

Search completed: June 27, 2002, 21:58:29
Job time: 662 sec

```

THIS PAGE BLANK (USPTO)